An Evaluation Approach for Assessing Program Performance from the State Energy Program

By
Nick Hall, TecMarket Works
Paul De Cotis, New York State Energy Research Development Authority
Marty Kushler, Evaluation Consultant
Lori Megdal, Megdal & Associates
Edward Vine, Evaluation Consultant

Introduction

This white paper was prepared for the Office of Weatherization and Intergovernmental Programs (OWIP) within the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy. The paper was written as a result of a request from OWIP for expert program evaluation advice and a proposed evaluation approach for assessing the State Energy Program (SEP).

In responding to this request the authors of this paper discuss three evaluation-related topics:

1. Evaluation's role in the process of assessing program impacts and helping to guide public policy decisions,
2. Some of the key evaluation policy and planning issues that the SEP program needs to consider before determining what to evaluate and before selecting the type of evaluation needed or the evaluation approach to be used,
3. A recommended approach for conducting evaluations of the national SEP program, with the wide range of energy efficiency and renewable energy programs implemented by the States through their SEP grant.

The primary purpose of this white paper is to develop a program evaluation approach for conducting an energy impact evaluation of SEP and the program measures carried out by the States, beginning in 2007 and later years. This timing allows for the development of this white paper, the attainment of an agreement within OWIP on the evaluation approach from which to move forward, and the subsequent implementation of that approach in time to document the effects from the SEP program measures implemented in 2007.

In conducting this work, the authors were guided by the principle that the primary objective of evaluation in this case, is to provide information to policy makers that documents the effects that SEP program measures are having in the energy efficiency and renewable energy markets in which they operate. If the State SEP activities are successfully achieving their desired effects, the evaluation can document these achievements. If not, the evaluation can help identify changes that can be implemented in order for the program to become more effective.

Of particular importance to the SEP evaluation effort is the fact that there are over 400 individual energy efficiency and renewable energy program measures/activities that are funded in whole or in part with SEP funds. These activities have a wide range of objectives, some of which are to

save energy, and some of which are to influence energy markets in a way that allows the markets to become energy efficient or to produce additional energy resources. That is, not all individual programs have objectives that can be indentified in terms of direct energy impacts (i.e. therms of natural gas saved, kW or kWh saved or produced, barrels of oil not needed, etc).

Many of the programs funded by SEP are market transformation, market preparation, or market development programs that have goals that are not related directly to saving or producing a given amount of energy, but instead are designed to allow other programs or market forces to achieve savings or produce new energy resources. Thus, the evaluations conducted need to focus on the goals and objectives of the individual programs rather than on a single objective that applies to all programs.

To determine the total impacts or effects of all SEP programs is a complex evaluation challenge. It would require evaluators to understand the various objectives of the total set of programs and structure evaluations to assess program performance toward those objectives. In order to make evaluation a manageable task, particularly given the relatively small budgets allocated to program evaluation, it is up to OWIP, and particularly the SEP team, to clarify the most important program objectives, and identify the highest priority data and information needed for program design, effects and funding determinations. This clarification will allow the evaluation efforts to focus on documenting the level of accomplishments of the program's objectives and assess if the attainment of those objectives was achieved in an effective and cost-efficient manner. The evaluations would also determine if the accomplishments are consistent with the purpose for which the program's funding was approved. This early identification of key evaluation data priorities will also help coordinate and tighten the focus of SEP programs funded by DOE.

Evaluations are conducted for at least two key reasons: (1) they provide feedback to program designers and implementers that allow programs to be improved, and (2) they provide documentation on the efforts and accomplishments of a program to make sure public funds are well spent. We encourage OWIP, and DOE, to consider an evaluation paradigm that encompasses these broad objectives, in order for the evaluation effort to provide a complete and informative assessment of the SEP. However, the task for the authors of this paper is more narrowly defined. As a result, this paper focuses primarily on the most important quantitative objectives associated with assessing and documenting program energy impact- related accomplishments to feed a part of the public policy decision making process. However, just as importantly this paper provides guidance to OWIP on how such a program evaluation effort should be conducted.

The development of this white paper is the result of a series of meetings between the authors and key managers in OWIP. In these meetings, it was decided that in addition to the presentation of a recommended evaluation approach, this paper must also convey key evaluation concepts that need to be understood so that the evaluation efforts can achieve the evaluation objectives of not only OWIP but also be sensitive to the evaluation objectives of the individual State Energy Offices, and operate within an evaluation framework that is informed by the full potential that can be realized via the evaluation effort. From this perspective, this white paper is as much an

information document from which evaluation policy can be informed as it is a recommended evaluation approach.

To arrive at the information presented in this document, the authors exchanged a series of evaluation concept drafts and evaluation issue outlines that fed discussions within the author group and discussions between the authors and the OWIP SEP Program managers. The authors of this paper have significant experience within the evaluation industry and are also familiar with the types of programs associated with the SEP program.

Background

In June of 2005, the Oak Ridge National Laboratory (ORNL) released an evaluation of the SEP program[1]. This study documented that the SEP program was cost effective and achieved energy savings that were acquired at a cost less than the cost of electric generation from power plants. In designing the approach for the study, ORNL needed to match a significantly restricted evaluation budget with a methodology that could be conducted within the available budget. The objective of the ORNL study was to estimate the energy impacts associated with the vast majority of the SEP program measures so that the SEP managers would know if their programs, overall, were cost effective and providing added energy resources to the United States. The study approach was based on a review of published evaluation literature linked with a structured allocation of program energy impacts identified in the literature in a way that proportioned the impacts identified across the same types of programs implemented by the State Energy Offices. That is, the results of the predominantly non-SEP evaluation literature were applied to similar types of SEP programs to obtain a general estimate of the level of energy impacts. While ORNL suggested that a more rigorous approach would be more accurate, the resources available for the study were not sufficient to conduct a more rigorous evaluation on more than a very few programs. Conducting more rigorous studies on a few programs would have left the vast majority of SEP programs un-evaluated, opening an even greater area for concern with the approach taken. Because OWIP needed SEP program-wide savings estimates rather than estimates for individual programs and because the evaluation budget supported only a literature application approach, the literature-based approach was implemented via a joint OWIP-ORNL decision. While both OWIP and ORNL desired a more rigorous evaluation of the SEP funded programs, such an approach was not possible within the resources available, requiring ORNL to use the literature-based approach. The ORNL approach and methodology was appropriate for its intended use, given the severe resource constraints.

Following the completion of the ORNL study the Office of Energy Efficiency and Renewable Energy commissioned the management and accounting firm of Deloitte & Touche (Deloitte) to review the ORNL 2005 report and identify evaluation implementation issues that EERE should consider in future studies. Deloitte's review indicated that the study did not focus on key metrics (lifetime energy savings) and was not grounded in reliable enough approaches to instill confidence in the energy impact information. Deloitte also indicated that the evaluation was not prioritized to focus on *the most important, most costly or least well understood* programs. They suggested the following:

[1] [1] Martin Schweitzer, Bruce Tonn, Oak Ridge National Laboratory, *An Evaluation of State Energy Program Accomplishments*: 2002 Program Year, June 2005, http://naseo.org/sep/documents/SEP_study.pdf

> "*Focus on a few programs with greater time and resources; a more thorough and rigorous evaluation of individual programs can be performed, particularly for those programs that are deemed to be most important, most costly, or least well understood*."

As a result of the review by Deloitte, OWIP hired TecMarket Works, an independent energy program evaluation firm, to write a white paper recommending an evaluation approach that would address the issues identified by Deloitte and provide more reliable estimates of energy impacts. TecMarket Works was also asked to form an author team consisting of some of the most experienced evaluation professionals in the industry to serve as a resource and a peer review team for the white paper.

OWIP SEP managers wanted an SEP evaluation approach that not only was independently developed, but was subjected to a peer review and advice process. This document is the result of that effort and presents a recommended SEP evaluation approach. This paper also discusses several ancillary issues that impact the evaluation process and the decisions associated with the evaluation effort. The evaluation effort suggested in this white paper is designed to be consistent with Deloitte's recommendations, and presents a systematic approach for more rigorously evaluating the higher priority programs on which the evaluation efforts need to focus.

This paper is based on an assumption that evaluation resources will be made available that are consistent with Deloitte's findings and recommendations as well as the approach presented in this paper. For comparison, California, which is the only other governmental organization in the United States with a similar but somewhat smaller number of programs (also located in a single state rather than spread out over all states and territories), has allocated, via a California Public Utilities Commission Order, $150 million to evaluate about 270 programs over a three-year period. While this budget allows for reliable evaluations of the high priority programs in California portfolio of programs, it is not enough to reliably evaluate all 270 programs. OWIP finds itself in a similar but more extreme situation in that SEP funds about 400 different programs and measure efforts. Indeed, allocating an evaluation budget equal to the entire federal SEP budget would not be sufficient to reliably evaluate all of the current SEP programs. However, if the evaluation efforts focus on the *"most important, most costly, or least well understood"* programs, this effort can be accomplished within a significantly smaller budget.

From a professional evaluation perspective, the authors of this paper suggest that about five percent of the annual program implementation budget be set aside as the minimum evaluation budget for assessing "*those programs that are deemed to be most important, most costly, or least well understood*." This recommendation is based on 25 years of evaluating energy efficient and renewable energy programs and represents a level of expenditures that several states have established for their utility program. While state utility programs budgets are both lower and higher than this recommended level, the five percent level is considered reasonable by most evaluation professionals and by several state regulatory agencies. However, at this level it will be necessary to prioritize not only the programs that need to be evaluated, but also the types of evaluations to conduct. Conducting reliable energy impact evaluations of all SEP programs would consume an evaluation budget that is greater than the program implementation budget for

all states combined. However, if the evaluation effort is well planned, well structured and evaluation priorities are well set, this budget level can be strategically applied to evaluate key SEP programs, and provide for a general estimate of program impacts. In order to accomplish this, a variety of assessment methods will need to be matched to the need for evaluation rigor for programs that are not considered the highest priority programs.

Matching Evaluation Efforts to Evaluation Needs

As noted earlier, the SEP Office provides funding for a wide range of energy-related programs at the state level that focus on a wide range of desired effects. A review of SEP program records contained on WinSAGA (the SEP program management database) suggests that there are at least 20 categories of programs[2]. Within these categories are different types of programs with different program objectives. Not all programs have energy impact objectives such as the production of a specific amount of wind-generated electricity, or the saving of a specific amount of natural gas, fuel oil, gasoline or electricity. Many programs are established to support other energy efficiency or renewable energy programs or to help develop energy policy that has the potential to influence energy production or savings. Likewise, other SEP programs can be classified as market change programs. That is, they are implemented to influence the way a market operates so that the market can influence the amount of energy produced or the level of energy saved.

The types of market change programs noted above typically do not have energy impact objectives. As a result, evaluating these programs to document the energy impacts that they achieve makes little sense. However, evaluating these programs to document if the program is achieving the goals and objectives on which the program is based does make sense. Any evaluation effort needs to not only prioritize the programs for which evaluations will be conducted, but also has to prioritize the objectives on which the evaluations should focus. Limiting the purpose of the SEP evaluation effort to focus only on energy impact objectives ignores some of the primary accomplishments of the SEP programs that are not designed to directly provide new energy resources or achieve energy savings. The evaluation planning process needs to consider energy saved, new energy produced and the program accomplishments that make it possible for other programs or market interventions to have energy impacts beyond SEP. For example, some SEP programs are funded to provide state energy policy development support. These programs produce no direct savings, however as a result of these programs states are more informed about what energy policies should be adopted for their state and provide a portion of the grounding information and analysis on which public energy policy can be developed. The evaluation focus for these types of programs may be the identification of the policy issues and considerations informed by the SEP program and the resulting policies that have been developed as a result. Then the evaluation, if needed, can focus on the energy impacts associated with the implementation of those policies. While few state or national energy policies are formed only as a result of these types of programs, and these programs cannot typically take credit for all energy impacts obtained under their influence, the program's influence can be

[2] This estimate is not comprehensive or based on a detailed review of program materials; rather it is based on a general review of the program information found in WinSAGA for several key states. Prior to the detailed evaluation planning efforts a more detailed review of the available program information will be conducted in order to prioritize the evaluation efforts.

evaluated and the evaluation can identify the energy impacts associated with policy changes made in part as a result of the SEP programs.

Some SEP programs can be characterized as market transformation programs (also called market change or market effects programs). These programs are often not designed to directly acquire energy impacts, but are designed to influence how markets operate so that they can operate in a way that saves energy or produces new energy resources. For example, an SEP program might focus on providing support for assessing sites for the construction of wind turbine generation facilities. While the program would not directly acquire the wind energy generated at an approved site, the program may have been instrumental at gaining approval for the location of the facility. Other programs work with wholesalers and retailers to help make sure that energy efficient products are available in the local markets, providing a wider choice for customers and capturing a change in stocking or sales practices that then help acquire energy savings.

The accomplishments of these types of programs would show no *direct* energy impacts, however the study might show, for example, that the program's efforts were influential in approving a location for a renewable energy facility or for allowing a greater penetration in the market for energy efficient technologies. What is more difficult to determine is the degree of influence that the program had on gaining the approval for the renewable energy site and determining if the facility would not have been constructed anywhere without the program's assistance. Likewise, it is more difficult to determine if a store's practice in stocking different equipment would have occurred without the market effects program that was specifically designed to change stocking practices. However, these researchable issues can also be successfully incorporated into an evaluation effort. The California Energy Efficiency Evaluation Protocols[3] (California Protocols) has a chapter expressly devoted to designing evaluations of programs that have objectives other than the direct acquisition of energy savings or resources. This document can serve as a guide for the SEP evaluation planning efforts.

Just as the primary objectives of the SEP programs are different from program to program, the impacts of the program do not always lie only in energy or energy-related changes. For example, energy programs typically stimulate the growth of the local economy and increase the number of jobs available within that economy. Likewise, energy programs often make companies more profitable and allow companies to improve their competitiveness, expand their markets or reduce their operating costs. Energy programs typically result in substantial reductions in greenhouse gas emissions (GHG) by reducing the amount of fossil fuel that needs to be burned. These programs are often considered the most cost effective way to reduce greenhouse gas emissions. SEP programs can also influence the ability of a population to function within the society in a way that benefits the individual program participant, the population as a whole or even the general conditions under which people live. Many SEP programs are designed to impact energy imports and can influence the need for maintaining reliance on energy supplies imported from countries that may not hold the same economic, social or military priorities.

Good public policy is seldom based on only one program objective or one type of program result, but rather is informed by a more complete set of information about a program's impacts.

[3] California Public Utilities Commission, TecMarket Works Team, *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*, April 2006

The SEP program evaluation planning efforts should consider a wide spectrum of program effects on which the evaluations can be focused, and consider evaluation approaches that document energy effects as well as a range of other SEP program objectives and program results. EERE and SEP program managers may want to work with their evaluation contractors and review the range of potential researchable issues on which the evaluations should focus and set evaluation priorities for both energy and non-energy related evaluation metrics in order for DOE to have a more complete understanding of the effects of the SEP programs.

Other Evaluation Issues to be Considered

Independence

Independence is the hallmark of objective evaluation. For this reason the U.S. General Accounting Office (GAO) often requires management audits and evaluations be conducted by individuals who have no stake in the results of the evaluation findings. The Office of Management and Budget (OMB) also supports independent evaluation and encourages evaluations to be conducted by organizations that do not have an interest in the results of the studies. California requires that evaluation contractors be independent of the implementation function and bars firms who are implementation contractors from performing impact evaluations of energy programs. DOE has recently moved away from having ORNL conduct the national weatherization program evaluation (as they have in the past) partly because of the need for unquestioned objectivity and the appearance of non-conflicting evaluation structures.

Yet to require *total* independence greatly increases the cost of the evaluation effort by requiring that all data collected in an evaluation be collected by independent individuals or organizations. Under these conditions, the requirement for independence is taken to the extreme. Because there are over 400 different SEP programs across the 56 states and territories, it is not possible to conduct totally independent evaluations without allocating an evaluation budget that is greater than the total SEP program implementation budget. However, the information collected and analyzed within the evaluation effort needs to be independent and non-biased to the extent possible. The level of independence in the data collection efforts will need to be structured to match the need for information, the priority of the evaluation effort, the expected objectivity of the data available, the cost of acquiring independently collected information, and other considerations. While the SEP evaluation efforts should be managed and conducted by independent evaluation professionals under the direction of the OWIP SEP managers, these studies should make maximum use of state data when that data do not bias the evaluation findings. The data collection efforts will need to be structured to make use of the available data in WinSAGA and additional information available from the states themselves.

Because WinSAGA may not contain all of the information needed to prioritize the programs for the evaluation effort, or to structure evaluation approaches, it may be necessary to obtain additional program information from the states in the early phases of the evaluation planning process, and to support the evaluation efforts when they are launched. However, the evaluation contractors should always keep in mind that the overall goal of any evaluation is objective, reliable evaluation findings, regardless of the source of the data used to conduct that evaluation. It is up to the evaluation contactor to inform OWIP SEP managers when the data collected from the states or provided in WinSAGA is not considered comprehensive or accurate enough to

conduct an evaluation, and plan for additional data collection efforts as appropriate. This may require the need to obtain additional information from the organizations implementing the programs, and in some cases the information collection will place an additional cost on these agencies when that cost cannot be incorporated into the evaluation effort. Typically these additional efforts do not require significant costs and typically represent the sharing of information already collected by the agencies implementing the programs.

Purpose of SEP Evaluation

Evaluation is by its very nature a part of the SEP program management and oversight function. It is a tool that is used to help SEP managers and policy professionals determine if their programs are on the right course, are achieving their objectives, and are cost effectively accomplishing their intended purpose. Evaluation is a tool to help managers determine when to change a program, or identify what parts of a program need to be changed, or to confirm that the program is on track, heading in the right direction. The very lack of an evaluation function within an energy program organization is often cited as a negative finding during a management audit of the organization's operations. As a result, it is important for policy makers and oversight professionals to incorporate the evaluation function within the program oversight structure and use the evaluation tools wisely to provide the information needed to support their management and oversight responsibilities. The use of the evaluation function should focus on the purpose of the programs that are being evaluated. As noted earlier, many SEP efforts are not designed to directly save energy, but are designed to support other efforts that do so or programs that build market capability to acquire energy resources. Still others are designed to support legislative or policy changes to help grow or protect the renewable or energy efficient opportunities in a state. It is up to the evaluation function to understand all of the objectives on which the studies need to focus. For example, a program that saves energy, improves profits and improves the economic stability of families may be more important than a program that only saves kilowatt hours. Likewise, a program that increases the economic health of a state and improves the competitiveness of the state's businesses may be more important than a program that saves energy but results in the loss of jobs. A program may be saving energy or be cost effective at providing new energy resources, but if it also harms the health of people impacted by the program then it may not be the best program to offer from a public policy perspective. A program that reduces the amount of gasoline used on America's highways may be good for the country, but if the program also results in increased shipping costs, higher consumer prices, and a reduced standard of living, then it cannot be considered successful even if the primary objective is achieved. There are always trade-offs that need to be considered in evaluation of government programs and policies.

It is up to the evaluation professionals in cooperation and coordination with the SEP program managers and oversight authorities to design evaluations that provide the information needed to determine if the programs are successful from an oversight and public policy perspective. Some of the researchable issues that SEP managers can consider in the evaluation planning process include:

- Is the program changing the market in a way that improves energy efficiency, or that increases energy supplies?
- Is the market being changed in a sustainable way?

- Are participants taking the actions recommended by the program?
- Is the program increasing the standard of living for participants and non-participants?
- Are the programs serving all customer classes and market segments?
- Are there specific programs or categories of programs that are more effective?
- Are there better ways to market a program?
- What fuel is being saved by the program?
- What is the level of carbon emissions that are being reduced?
- What is the level of net-adjusted impacts associated with the program?
- Is the program resulting in spillover savings (see below), and at what level?
- Are programs taking advantage of leveraging potential?
- How much energy is being saved?
- What is the load shape of the energy saved?
- Is the renewable energy being effectively dispatched and used to offset fossil fuel-based generation?
- Are the programs affecting the health and safety of participants or non-participants?
- Are the programs resulting in economic development within the communities in which they are offered?
- Are the savings benefits resulting in less stressful economic pressures on participating households and families?
- Are the programs helping to make the country more secure or less dependent on energy imports?
- Has the SEP program enabled states to offer programs that they would not have offered without the SEP financial support?
- Has the SEP program resulted in greater capacity within the states to deal with or solve energy related issues or to develop more effective energy policy because of the SEP support?
- How much carbon dioxide is not being released into the environment?

The above list represents a non-comprehensive sample of the researchable questions that can be addressed within the SEP evaluation efforts. The SEP planning process should consider a wider range of researchable issues than the gross levels of energy saved or the gross levels of energy generated. This consideration should not only identify the priority programs, but also the priority of the researchable issues that can be addressed via those evaluations. Once both the program priorities and the researchable issue priorities are established, the evaluation efforts needed to address those priorities can be planned, budgeted and conducted.

Types of Evaluation[4]

There are two basic types of evaluation that are used in energy efficiency program evaluation field. These two broad categories of evaluation are:

1. Effects Evaluation (energy and non-energy effects), and
2. Process Evaluation.

[4] Material on types of evaluations modified from the text of the California Evaluation Framework, TecMarket Works, June 2005.

Effects evaluation includes many different types of program "effects" studies. These include the energy and demand savings that are the focus of impact evaluations as well as other studies that document the short-term, mid-term or long-term effects of a program. Energy impact (effects) evaluations comes in many forms, ranging from simple engineering estimates applied to counts of measures installed, to sophisticated statistical models incorporating measured energy consumption data and numerous other variables and statistical corrections. As noted earlier, other types of programs can have different types of effects evaluations, such as market effects and market change studies associated with market transformation programs. Information and education programs also use effects evaluations to document the changes that occur as a result of the program's efforts. Effects evaluation also increasingly has come to include important elements that go beyond the effects that are the primary program goals to secondary effects to "non-energy effects." For example, energy efficiency programs are considered to be the most cost effective way to reduce greenhouse gas emissions (GHG). The one unifying theme in all these activities is the fundamental purpose of effects evaluation: *to serve the key function of documenting and measuring effects caused by the program.*

Similarly, process evaluations can be accomplished using a variety of methodological techniques, but its purpose is to serve the key function of documenting and understanding how the program is operating and to identify opportunities for program improvement.

An expanded view of process evaluation can include studies to understand market operations and processes and how market transformation programs are operating within those market processes. In this way, the process and effects components of evaluation can work together to provide both the functions of *effects* and *process* evaluation for market transformation programs. It is possible to further delineate several sub-categories of market transformation program evaluation, each with a somewhat different focus and purpose. In this regard, the studies would distinguish:

1. Market effects evaluations, which examine the changes in a market caused by a program. Many SEP programs are designed to achieve those types of effects.
2. Market operations and baseline evaluations,[5] which look at how a market operates, the key information points, information hubs, and how products flow. This type of evaluation is used by evaluators for assessing program performance or program designs and operations.
3. Marketing evaluations, which examine the effectiveness of a set of marketing or market outreach efforts.

In many cases, a well-designed and comprehensive evaluation approach will incorporate both of the basic types of evaluation (effects and process) to provide a complete assessment of the program, so that program and policy decisions can be grounded in the rigorous and objective information needed to support a program or policy decision.

What this means for the SEP program evaluation effort is that the evaluation objectives associated with the highest priority programs will need to be identified prior to selecting the type

[5] Sometimes referred to as "market characterization."

of evaluation to launch prior to developing the detailed evaluation approach. SEP funds several market transformation programs that are designed to change how a market operates, typically to increase the use of high efficiency products (ENERGY STAR refrigerators), to change the behavior patterns associated with something that uses or affects energy consumption (turn down thermostats), or to place additional generation resources (renewable energy facilities) into the market. The early evaluation planning efforts will need to focus not only on what programs are high priorities, but also which researchable questions are to guide the evaluation effort.

Budgeting for Evaluation

As discussed earlier, the budget for evaluation must be sufficient to achieve the overall purpose of the evaluation. As noted by Deloitte, the ORNL study did not provide information needed on the effective useful life of the achieved energy savings, did not address several important evaluation metrics and was grounded on an evaluation approach that was considered somewhat unreliable. Yet the ORNL study provided value for the evaluation budget by providing an indication of the level of savings SEP programs have achieved by relying on the evaluation literature for documenting savings at the program level. This approach is often used in the evaluation industry when the evaluation budget does not support primary research.

Budget estimates for the evaluation efforts typically range from a high of ten percent of the program budget to a low of about two percent. There are numerous examples of budgets set at higher and lower levels. Nationally, about four to six percent of energy efficiency program budgets are spent on program evaluation. Evaluation budgets are often determined by the number of programs that need to be evaluated each year and the reliability stakeholders need in the evaluation findings. The Deloitte assessment of the ORNL study suggested that SEP evaluation budgets be allocated at a level needed to reach more reliable conclusions. The same recommendation was made for California's energy efficiency programs when the California PUC needed more reliable evaluation results than could be achieved with a 4.25 percent evaluation budget. The PUC ordered that evaluation studies be funded at eight percent of the program budget in order to obtain more reliable findings.

The most important consideration for establishing evaluation budgets is an understanding of the reliability needs across the researchable issues to be addressed in the study. If OWIP SEP managers and program oversight stakeholders do not need a high degree of statistical precision when evaluating SEP, then the budgets can be lower. If the stakeholders need relatively more precise data for public policy decisions, then the budgets need to be set at levels consistent with the researchable methods employed to deliver that level of reliability. One of the most important reasons for evaluation is to improve the operations and cost-effectiveness of energy programs. Evaluation budgets that are set too low to accomplish these goals are of little value to any stakeholder. However, evaluation budgets that are set higher than needed for these goals may be a waste of valuable resources.

There is no single specific percentage of a program's budget that should be allocated for the evaluation process. Evaluation budgets for small pilot programs (less information available), that are testing new designs or delivery concepts, may need to be set at a level higher than the program's costs in order to collect and analyze the information needed to determine if the program should be continued or expanded. At the other end of the scale are larger programs

offering technologies and services that have been evaluated in the past, offered in well-understood environments in which the energy impacts and operational procedures are well-documented, and in which the procedures are working smoothly and efficiently. In these cases, a $4,000,000 program may need only two or three percent of the program budget to periodically verify energy impact. Most energy program evaluation budgets can be set somewhere between these two extremes.

At the end of the day, and after all arguments have been aired, policy makers still need to allocate a budget to the SEP evaluation effort, and that budget has to be set at a level that supports the needed study results. Adequate evaluation budgets are generally lower than ten percent but higher than three or four percent of the program implementation budget. Budgets can stay toward the lower end of this range when prior evaluations are drawn upon and circumstances are explored to determine the necessity for repeat evaluations. The SEP program budget is about $45 million a year. If SEP were to allocate a budget at the lower end of the above scale, at about 4 or 5 percent of the program budget, the SEP evaluation budget can be set at about $1.8 to 2.0 million per year. If reliability is an issue, as noted in the Deloitte assessment, the budget can be set near the 10 percent level, at about $4.5 million per year. However, if priorities are set so that higher priority programs receive more reliable evaluations and lower priority programs receive little or no evaluation (as suggested in the Deloitte assessment), the evaluation budget can be set near the lower end of the scale and focus on key high priority programs and rely on estimates of impacts for the lower priority programs. It is also possible that a larger evaluation budget can be established in the earlier years of the evaluation effort, with budgets decreasing over time as confidence and reliability of evaluations increase and as the SEP Office gains experience with and uses the evaluation findings.

Evaluation Metrics

The Deloitte assessment of the ORNL report identified a key energy impact metric that was not included in the ORNL study. This metric is the lifetime (also call the effective useful life or lifecycle) energy impacts. Unfortunately, most of the pre-2005 evaluations within the energy program field focused on annual energy impacts associated with the first full year of program operations. Few studies prior to 2005 focused on lifecycle savings. As a result, the ORNL report was unable to report estimates for lifecycle savings using a literature based approach. Recently more studies have focused on lifecycle savings, or what is more commonly referred to as effective useful life (EUL) savings. EUL savings are the energy impacts that will occur each year over the lifecycle of the events caused by the program. Lifetime or lifecycle energy savings are an extremely important consideration when thinking about energy impact documentation objectives of the evaluation effort. While most evaluations are not designed to assess the exact number of years over which savings are achieved[6], the evaluation studies should estimate the energy impacts that are expected to result each year over the effective useful life of the changes caused by the program. This means that in addition to the estimation of the lifetime energy impacts, the evaluation effort will also have to find a way to identify the effective useful life of those impacts. However, there are considerable extra costs associated with attempting to document lifetime energy impacts. Moreover, as noted above in the section about the purpose of

[6] Studies that document the number of years over which energy savings are achieved are typically market-sector based studies (residential, commercial, industrial) that examine the length of time an energy savings technology is used in an operational environment that provides the savings.

evaluation, there are a great number of other reporting metrics that need to be considered. Each of the sample researchable issues identified in the list presented above has a set of reporting metrics that apply to that issue. While the energy impact metrics are the more important metrics to be produced from the evaluation, OWIP SEP managers may want to consider a wider range of researchable issues on which the evaluation should focus, and incorporate additional metrics into the evaluation plan and the reporting requirements. These metrics can be quantitative as in the kWh saved, the demand (kW) reduced, the gallons of gasoline saved, or other such metrics. However, there are also the more qualitative metrics to consider. For example, studies of weatherization programs have focused on improvements in a family's disposable income, changes in comfort levels, fewer sicknesses, and fewer days lost at work. Other studies have focused on the number and type of jobs brought into a state and the net economic impacts of a program to a state. While other studies have focused on the return on investment from energy programs where for each dollar spent, x number of dollars are moved through the local economy, stimulating growth within that economy. In determining the evaluation metrics on which a program should focus, SEP managers may wish to incorporate the needs or desires of the state energy offices that operate the programs to determine which additional information the evaluation can provide that are needed by the states offering the programs. Because one of the primary jobs of the evaluation effort is to improve programs, incorporating the needs of the state implementers can help improve programs.

The Importance of Program Tracking Data

In their assessment of the ORNL study, Deloitte concluded "*Given limited resources and insufficient data from the states, the SEP Report's failure to provide this type of assessment was almost inevitable".*

In addition to having adequate evaluation budgets to conduct the needed studies, detailed program information must also be available from which evaluation efforts can be planned. The availability of detailed information about the intended effects of a program and the operational approaches used by that program to achieve the program's intent is critically important for developing research methods to document the end-effects accomplishments. Deloitte's assessment is correct in that both adequate evaluation resources must be made available to document effects, and detailed program operational information must set the groundwork for developing the research methodologies on which a study is based.

In fairness, the June 2005 ORNL referenced by Deloitte & Touche was severely under-funded for the purpose for which the Deloitte assessment assumed. The ORNL study also suffered from the lack of a comprehensive database detailing the program descriptions and operational theories on which each SEP program is based. Likewise, the database available for the ORNL study did not present the detailed efforts and activities associated with each of the program's efforts in a way that allowed the ORNL analysts to fully understand the range of activities conducted, the market served by the program or the services offered[7]. Without detailed state-program-specific program operations, activities and accomplishment databases the ORNL analysts did not have adequate information on which their energy impact estimates could be adjusted. The absence of

[7] Establishing such a database of program efforts, markets and activities would have added a substantial cost to the ORNL assessment because few states have established these types of databases that are typically common to utility, public goods or public benefits charge funded programs.

both detailed information and an adequate budget meant that ORNL had to find the best approach possible under the significant restrictions associated with the study.

Within these constraints, the ORNL report is among the best evaluations that can be expected from even the most skilled evaluation professionals. However, this does not mean that the OWIP SEP managers should give up on finding a more rigorous or more reliable evaluation approach. Rather, it means that they are going to need to budget evaluation resources that are consistent with the reliability needs of the study. However, just as importantly, it also means that the evaluation effort must spend time and resources investigating SEP programs to the degree that a reliable evaluation approach can be designed. Deloitte suggested that the evaluation focus on "*the most important, most costly or least well understood*" programs. However, without good program information, it is difficult to know which programs are the most important, most costly, or the least well understood.

As noted earlier, the DOE maintains a program tracking database (WinSAGA) that supports the SEP managers' needs with regards to program funding and SEP's historic oversight efforts. A review of WinSAGA indicates that some states provide detailed program-specific information about their programs, the funding sources and funding levels. However, other states provide less data, making it difficult to understand the programs enough to structure rigorous evaluation approaches. Because the funding levels for SEP are not adequate to develop and maintain detailed program tracking databases in each state, the requirement for detailed program information to support the evaluation efforts amounts to an unfunded mandate, if it were to be required. As a result, the SEP program office is forced to conduct evaluation efforts with the data they have, rather than the data they need.

Thus, one of the first things that will need to be done before evaluations can be launched is to organize the WinSAGA data into program-specific evaluation planning formats. That information will need to be supplemented with additional data collection efforts to more fully understand the level of importance of the programs, the dollars spent (total and leveraged funding) on the program, and the acquisition of enough program information that a the program can be classified into an evaluation cluster for planning the studies, and prioritized for assigning rigor levels associated with the priorities of the evaluation effort. The lesson learned from this is that without evaluation-structured program tracking databases the evaluation efforts are hampered from the beginning. Therefore, there is a need to spend efforts acquiring the data that Deloitte suggests is needed to prioritize the evaluation efforts. We agree with Deloitte in that the studies themselves need to focus on *the most important, the most costly or the least well-understood programs*. The evaluation approach developed in this white paper is specifically designed to do exactly that, once there is enough program-specific information on which those priority decisions can be made. However, OWIP may need to determine if WinSAGA should be restructured to collect more of the necessary information from which the prioritization of evaluation efforts can be made in order to streamline the evaluation planning process in the future. Another approach for obtaining evaluation planning information (rather than restructuring WinSAGA) is to place this data collection activity within the early evaluation planning efforts (by interviewing state program managers) and repeat this effort only when an up-dated evaluation planning effort is required. For some states, this will mean little or no additional information collection. For other states, this will mean that more information about

program intent, timelines, funding, funding sources, operational practices, stakeholders, stakeholder involvement, projected end-effects and other metrics may be needed.

Causal Effect Analysis

An important component for most impact evaluations is estimating the program-induced effects, or the proportion of effects seen that are caused by the program. These analyses are guided by the cause-and-effect scientific literature. Measuring the program-induced impacts can be accomplished through research design, direct inquiry or specialized studies designed for this purpose. The three primary requirements for causal attribution are: association (correlation) with effects, temporal precedence (effect occurs after the intervention) and other potential causes ruled/measured out (no spurious causal correlation). The most difficult of these is the last one, ensuring that the program's measured net effect is due to the program and no other potential causes. Research design can be used to accomplish this by ruling out potential threats to validity.[8]

Most evaluation studies estimate program-induced effects or the proportion of the impacts that are caused by the program. Essentially, a study's causal effect analysis attempts to attribute the effects of the program with the causes of those effects. There are many research designs and statistical methods that can be used to unearth the program-induced effects. The causal effect analyses are attempting to determine what would have occurred in the absence of the program intervention.

For example, if a participant was going to take the actions recommended by a program even if the program was not offered, then it is difficult to give full credit to the program for causing the actions taken. It is likely that for these individuals the program may have had no impact on the actions taken. This type of attribution adjustment in energy efficiency program evaluation is typically referred to as the "free rider" adjustment. This is one of three types of program effects examined for direct participation programs using direct measurement methods (as opposed to market-based evaluations or other effect evaluation designs). The program's effects are discounted for participants who have received a "free ride," so that net energy savings are less than the total (gross) energy impacts achieved by the program. To the extent that the program caused the participant to take the program recommended (or the incented) action, the program obtains credit for that action in its estimate of net energy impacts. The savings from the actions that would have been taken in the absence of the program are referred to as the naturally occurring energy impacts or the impact occurring because of non-program causes.

A second type of direct program impact arises when trying to recognize the achievements of program participants who invested in energy efficiency or renewable energy generation outside of their direct participation in a program but due to their program participation. For example, a program participant receives incentives from a state energy office to buy three compact fluorescent light bulbs (CFLs). The participant likes the savings from the CFLs and goes out and buys five more CFLs without incentives. Should the SEP program, for example, take credit for the additional savings from these five CFLs that were not directly induced by the program via an

[8] A classic textbook on research design, and the threats to validity that should be addressed, is: Thomas D. Cook and Donald T. Campbell. 1979. *Quasi-Experimentation: Design & Analysis Issues for Field Settings,* Houghton Mifflin Company: Boston.

incentive? This type of attribution is referred to as a "spillover" adjustment. That is, it focuses on identifying the amount of impact that is achieved by the "spilling over" of the program's effects from participants outside of the program per se ("participant spillover").

California's evaluation protocol provides three different levels of rigor for evaluating free riders and participant spillover: (1) the basic rigor level relies on participant self reports; (2) the standard rigor level has three options (a) participant and non-participant analysis of utility consumption data that addresses the issue of self-selection, (b) enhanced self-report method using other data sources, and (c) econometric or discrete choice analysis with participant and non-participant comparison addressing the issue of self-selection; and (3) the enhanced rigor level using triangulation where more than one of the methods in the standard rigor level are used.

A third type of direct program impacts is that which arises when trying to recognize the achievements caused by the program when the program was not directly involved in obtaining that effect. An example of this indirect effect is when a participant tells their neighbor (a non-participant) about what they have learned or achieved as a result of a program in a way that causes the neighbor to take a similar action without participating in the program. Similarly, a mid-market actor could participate in providing services for a direct program and learn about business and customer benefits of promoting energy efficiency. Then they stock more energy efficiency models or promote them more. These types of impacts are referred to as a "spillover" or "free driver" (in the case of impacts on mid-market actor behaviors that create energy savings in their customers due to their stocking or promotion, driving their customers to achieve energy savings). That is, it focuses on identifying the amount of impact that is achieved by the "spillover" of the program's effects from participants to non-participants.

There are two primary approaches for estimating causal attribution for non-participant spillover: one uses a preponderance of evidence approach and the other uses a modeling approach. In the former, the evaluator uses multiple data sources to draw conclusions about the presence and attribution of non-participant spillover (through interviewing and surveying knowledgeable market actors). There are two levels of rigor associated with this approach: (1) the basic rigor level relies on a representative sample of market actors to provide self-reports on perceived spillover, and (2) the enhanced rigor level uses quasi-experimental or experimental design with comparison groups to evaluate spillover. In the modeling approach, the evaluator uses multivariate models to estimate net market effects, such as non-participant spillover.

The energy efficiency program evaluation literature is rich in examples of different approaches for measuring free riders and spillover, and the reader is referred to the references listed at the end of this white paper for further information. There is also an even larger body of literature on estimating net effects for many other types of programs. For the purposes of evaluating the SEP Program, evaluators will need to determine which programs need to have free rider and spillover adjustments and develop the approaches for applying these adjustments at the program level or develop research designs to derive net impacts. These considerations would be a part of the more detailed program-specific evaluation approaches that would be developed and approved by OWIP SEP managers prior to the launching of the program studies.

One of the more difficult aspects of estimating net impacts is separating out one specific program's impacts when there are many programs and other influences on the action or market being examined. Though difficult, it is important that evaluators provide their best estimate of program impacts that are truly net of what would have occurred in the market or due to interventions by other market or policy actors. There may be circumstances when many public and/or private entities are involved to encourage or support energy usage changes that may make complete scientific attribution between them cost-prohibitive or even impossible. In these cases, the independent evaluators should at least provide a review of the other causal effects for the changes seen and a recommendation on how to attribute impacts to the program being evaluated versus that of the other contributors. The most common method for doing so has often been by proportion of the expenditures made for all those intervening in this market or action (see below).

Allocating savings between different intervening programs or entities is sometimes accomplished through negotiation between the parties rather than evaluator assessment. This is often a less desirable method than that done by an experienced independent evaluator because it is important that the final net program impact be estimated to be as free of as much potential bias as possible. It also should be developed to ensure trust in the impact estimate and cause allocation efforts. This can be more difficult to accomplish if there is potential conflict of interest or even perceived conflict of interest. Yet, there are circumstances where negotiated allocation of savings is the most defensible and cost-efficient method to arrive at a net program impact estimate. This method of attribution is most clearly referred to as "policy attribution".

Policy attribution is associated with recognizing accomplishments associated with the various program providers and sponsors. For example, if a program is funded by four funding sources and is supported by in-kind contributions from two other organizations, how much of the accomplishments of the program are attributable to the different sponsoring organizations? This is a complicated issue with which the evaluation community has struggled for years. There is concern that minor funding or support organizations may not be critical to the success of a program. However, discussions with program managers typically indicate that all contributions are critical to the success of their program. These managers tell evaluation professionals that without these contributions they would have to de-scale or reduce services, making the programs less effective. In this scenario, a small part of a program that is funded by a minor funding source may provide key benefits on which the program relies for its success. Alternatively, there may be funding sources or in-kind contributions that are ancillary to the primary focus of the program, but which may be critical to the program's success but may not be a significant part of the program from a budget or activities perspectives.

This attribution problem becomes an issue when multiple organizations are separately reporting on the same amount of energy savings or greenhouse gas reductions. For example, if a program has an impact that is counted by four organizations and these impacts are then totaled to track a state's total effects, the total may be many times the actual program accomplishments because of double counting of the same impacts. This issue is a growing concern as greenhouse gas credits for energy efficiency programs are being developed. It is highly possible for a company to participate in an energy program and have the greenhouse gas reductions (credits) claimed by the SEP program which funded a part of the program, claimed by the state that implemented the SEP

program, claimed by two other funding sources that supported the program, and claimed by the business taking the action. Sorting out all of these claims can make the attribution process crucial to the counting effort when there are conflicting claims for the attribution of the benefits.

With regards to the SEP evaluation efforts, there is a need to develop attribution approaches that can be successfully implemented and at the same time reflect an acceptable level of reliability in the results. From this perspective, most organizations elect to use the proportion of the program's funding source as the primary (or only) metric for attributing savings to program sponsors. For example, if SEP provides 50% of the program's funding, then SEP would receive 50% of the energy impact credits associated with the program. However, even this approach is problematic. If a non-SEP program provides services that speed the approval process for constructing a renewable energy wind facility, but is not involved in the actual design, construction, commissioning, and operational efforts, what portion of the resulting power is a function of the SEP program? The answer can be anywhere from zero percent, to all of the credit because without the expediting efforts of the program the facility investors may not have gone forward with the facility.

In this white paper, we do not propose to solve the policy attribution of program effects. The purpose of this section of the white paper is to bring attention to the causal effect analysis and attribution issues, so that it can be incorporated into the detailed planning process associated with creating the detailed evaluation design for the SEP program evaluation efforts. The reason we are moving this issue into the program evaluation planning process is that the best place to deal with these issues is at the program level (*i.e.*, it is more than a technical evaluation issue – it is also a policy issue that needs to be discussed by evaluation managers, program managers and policy makers). As a result, the detailed program evaluation planning process must include an approach for estimating causal effects and attributing savings, and that approach must be informed by SEP policy considerations.

Need for Program-Specific Impact Planning Data

One of the objectives that OWIP SEP managers should consider in establishing a routinized evaluation effort within the SEP is the need for a set of energy impact estimates for different types of programs within different climate zones. At this time, there is no baseline that states can use to set as the typical energy savings that can be achieved from a program. As a result, SEP managers are left to rely on the estimates of impact provided by each state without a good understanding of the approach used to estimate energy impacts. If the evaluation effort were to build a program-specific database that recorded the results of the SEP evaluation efforts for different types of programs, then one can develop an adjustment approach for increasing or decreasing the savings as a result of specific program and climate conditions. As noted below, SEP managers would not need to rely as much on expensive, comprehensive evaluation studies once the evaluations were conducted to establish the saving levels and construct an adjustment approach. This is not something that can be completed in a couple of years, but will require several years of evaluation efforts, so that the data can be constructed over time, consistent with the completion rate of the evaluation efforts. If the evaluations provide only the evaluation results, then the results for a specific program or set of programs at a given point in time can be well understood. However, if the evaluation studies were used to build a program-specific energy impact estimation database, then SEP managers could use that database to report

expected program savings as the states' SEP plans are submitted, or states could use the database to estimate their program's impacts. Because the database would be grounded in completed evaluation results, the projected savings would be more accurate than relying on estimated savings provided by each individual state.

As this approach is developed, the cost of the evaluation efforts can be reduced as more SEP managers rely on the increasing accuracy and comprehensiveness of the database. California uses a similar approach in that the evaluations completed over the last 12 years have been used to build what is called the Database for Energy Efficient Resources (DEER). The DEER database is adjusted periodically as new evaluations are delivered. Program designers, managers and oversight professionals use that database to estimate program impacts with increasing levels of accuracy. Because the DEER database is grounded in newly completed program evaluation results, it grows progressively more and more accurate each year. SEP managers should consider the potential benefits of developing a database for SEP programs and use that database as the platform from which energy savings can be estimated rather than funding new comprehensive evaluations of programs that may not need them because the database would already have an impact estimate for that program under the program's operational conditions offered in a specific state or climate zone. If this database were to be constructed, the builders of that database could rely on SEP evaluations and on evaluations of SEP type programs. It is important to note that the database needs to be supported annually; otherwise, the value of the database will diminish over time.

Boundaries of the Effort

The evaluation approach developed for the SEP Program needs to be adaptable so that it can cover any type of SEP program measure offered within the 56 states and territories. These programs cover a wide range of program types and service offerings. There are over 400 different types of programs funded, in part or in whole, by the SEP. These programs need to be considered in the evaluation effort priority setting process. However, in developing this white paper, we needed to set the boundaries associated with the development of the evaluation approach presented. The underlying boundaries for the evaluation approach are as follows:

1. The approach must be flexible, so that it can include all program types funded by SEP dollars.
2. The approach must be structured within an annual $1.8 to $2 million evaluation budget that may change from year to year, with specific evaluation studies planned for each year over a four or five-year evaluation cycle totaling $7.2 to $10 million. After the completion of the evaluation cycle, the *most important, most costly, or least well-understood* programs will need to be evaluated on a continuing basis at a level of rigor acceptable to SEP managers and policy makers.
3. The approach must include the creation of a database that includes the results from previous evaluation efforts.
4. The approach must provide evaluation options, so that the methodologies selected best match the research priorities and funding constraints on an annual basis.
5. The approach must reflect SEP evaluation priorities (that can change from year to year).
6. The evaluation approach must incorporate a focus on energy savings and renewable energy generation as key components of the evaluation objectives.

7. The evaluation approach must also be able to focus on additional researchable issues as needed, which can be incorporated into the impact evaluations (including both energy-related quantitative metrics such as GHG emission reductions, as well as broader program objectives where appropriate, such as market effects).
8. The approach must make use of program descriptive and program assessment information/metrics collected and maintained by SEP. New data collection required from the states needs to be set at the minimum necessary level to understand the programs to be evaluated and to successfully complete the studies.
9. The approach must be supported by SEP managers and peer reviewed by professional evaluation experts, so that the findings are reliable and can be trusted to the extent possible within the resources available.

The Evaluation Approach

The remaining sections of this document present an approach for structuring the SEP evaluation efforts, so that evaluation priorities can be identified, enabling the development of an evaluation planning effort that focuses evaluation resources on the highest priority programs. This approach is based on the assumption that the recommended evaluation budget discussed earlier will be allocated to the evaluation effort.

SEP Program Characterization

The first step in the evaluation effort is to understand the SEP programs offered and to sort those programs into evaluation groups for prioritizing the research efforts.

There are over 400 individual SEP programs included in the USDOE-SEP program database. These programs fall within 13 high-level program classifications. The high level classifications include the following types of programs:

1. Loans, grants, and incentives
2. Building retrofits
3. Government, school, institutional procurement
4. Renewable energy market development
5. Traffic signals
6. Workshops, training and education
7. Energy audits (residential, commercial, industrial, agricultural)
8. Technical assistance to building owners
9. Building codes and standards development
10. Energy efficiency rating and labeling
11. Tax credits for energy efficiency and renewables
12. Clean energy policy support
13. Other programs of various types and configurations

Within these classifications, there are many different types of programs. Essentially, there are few programs that are alike in every way; however, many programs are alike in very general ways. For example, there are many types of renewable energy development programs. These program all focus on some aspect of the renewable energy industry, but can be very different in the services provided. These programs range from providing help in locating and permitting

renewable power facilities, to renewable technology information programs, to advice on purchase decisions, to programs that increase market demand for renewable energy. Similarly, there are different types of audit programs. These programs vary in the markets they serve, the types of facilities they cover, and the service associated with the audit. For this white paper, we did not review the SEP programs funded to determine what types of programs reside within the 13 different classifications. However, a thorough review of the programs and the establishment of program evaluation groups will need to be conducted as part of the early evaluation planning efforts. Because of the great diversity of SEP programs, we expect that we will need to establish program evaluation groups. These groups will consist of the different types of programs that have enough similarity that they can be evaluated as a cluster of programs, even though the evaluation approach may differ across the individual programs within a group. Because of the need to prioritize programs within the limits of the evaluation budget, not all programs within a group will be evaluated. When programs are similar enough to evaluate as a group, it may be necessary to sample programs from that group instead of evaluating them all, depending on the evaluation priorities.

In order to establish the evaluation groups, a program characterization review must be conducted at the program level. This review will sort the SEP programs into evaluation groups based upon the individual program characteristics found in the WinSAGA database. WinSAGA contains program descriptions of every SEP program; however, as noted earlier, some of the descriptions are more detailed than others. For most programs, the characterization can be conducted using WinSAGA populated information, but there will be a need to gather additional program information for selected programs that are not as fully described in WinSAGA. This effort must also be budgeted into the evaluation planning process so that the evaluation priority setting assessment needed to support the evaluation planning effort is grounded on reliable program information. As a result, one of the first efforts associated with the evaluation planning process will be to acquire the information needed to group programs into evaluation clusters and assess and prioritize the SEP programs. The results from this effort will be the development of a set of program evaluation groups with a description of the characteristics that make the group suitable for grouping and descriptive information about the characteristics of each program that need to feed the efforts for prioritizing programs within an evaluation group. These data will include the budget for the program, the SEP contribution to the program, the expected range of effects from the programs as projected from the ORNL study for that type of program, and other information as appropriate. While the Deloitte review did not find that the program savings in the ORNL report can be considered reliable, the ORNL estimates provide a proxy for the general level of savings that may be expected, so that program groups and programs within a group can be compared by their expected energy impacts. In addition, WinSAGA also has information on some programs that provide expected energy savings information. Likewise, the evaluation literature also contains information on which expected savings can be obtained and used to support the evaluation prioritization efforts.

The results from this program characterization effort will feed the program evaluation priority assignments that lead to the identification of which programs to evaluate each year during the multi-year SEP evaluation cycle. For the characterization efforts to be successful, we will need to know enough about the programs to have a general understanding of the program theory supporting the program's operation and the activities undertaken by the program to achieve the

expected results. Because many programs do not fully describe the expected end result of their program's activities, the data collection effort will also need to identify the end results in more than just general terms. That is, the expected results will need to detail what it is that the programs are specifically designed to accomplish to the extent possible. As a result, it will be important to involve key SEP managers in each step of the characterization and prioritization efforts to make sure these efforts reflect the way the programs are operated and to accurately capture the services provided. As noted above, it may also be necessary to contact state program managers to confirm the interpretations of the WinSAGA data or to collect additional information about a program to allow its classification within a program evaluation priority group.

Assigning the Evaluation Objectives

Once the programs have been sorted into their evaluation groups and linked with the program descriptions and the key program metrics associated with each program (total funding, SEP funding, estimated impacts, etc.), the evaluation objectives for each evaluation group will be developed. This effort will also need a supporting budget and will be conducted in conjunction with the SEP program managers, so that DOE's SEP evaluation issues can be developed, reviewed and approved for each evaluation group. While energy impacts will be the primary objective of most evaluations, the review by Deloitte correctly indicated that evaluations of programs for which there is little information available on the effects of the program may be more important than programs in which the effects are better understood. For this reason, it is necessary to review the evaluation groups and identify a range of possible evaluation objectives that are needed by the SEP program managers. As noted earlier in this paper, there may be a wide range of researchable issues important to SEP managers. A review and discussion of these researchable issues will enable the evaluation planning efforts to incorporate a range of evaluation objectives into the evaluation planning process. Typical evaluation objectives and the ones on which most studies will focus include SEP's primary energy savings objectives. These include:

- Gross and net reductions in natural gas use (mmcf),
- Gross and net reductions in heating oil consumed (gallons),
- Gross and net reductions in petroleum-based transportation fuel used (gallons),
- Gross and net reductions in electricity consumed (MWh),
- Gross and net reductions in electricity demand (MW),
- Gross and net increases in renewable energy installed electric generation capacity (MW), and
- Gross and net increases in renewable energy generated (MW).

However, in addition to the above energy effects objectives (metrics), SEP should consider other objectives that enable policy to be based on a broader scope of program effects. For example, many of the SEP programs are focusing on market transformation effects, yet SEP evaluations and reporting metrics have yet to establish market-based change metrics associated with these market transformation programs, even though substantial funding goes into market change program efforts. While all SEP programs focus on energy saving or generating/supplying new energy resources, market transformation programs accomplish this objective via longer-term changes to the operations of a market. As a result, evaluation objectives need to focus on how

the program is changing the operations of the market and if these changes are producing the desired energy effects.

Many states that offer SEP programs also have non-energy related objectives. For example, states that offer programs through their economic development offices instead of offices dedicated to only energy savings are as much focused on job creation as they are on energy impacts. Where there are objectives in addition to energy savings, these expected results can be considered. Non-energy objectives that should be considered in addition to the energy-saving objectives, and which are linked to the researchable issues discussed earlier in this paper, include:

- program-induced changes to the way a market operates,
- sustainability of market change,
- financial impacts of the program to households and businesses,
- reduced demand for imported oil,
- reduction of carbon dioxide emissions,
- impact on the health or safety of the participants,
- jobs created or lost in the community,

In this step, the evaluation contractor will work with the SEP managers to identify the energy and non-energy evaluation objectives that can be associated with each of the evaluation groups identified earlier in the evaluation planning process. Not all of these objectives, especially the non-energy objectives, will survive to have a high priority placed upon that objective when program evaluation priorities are developed. However, the development of a more comprehensive set of potential objectives allows the SEP managers to consider all of the objectives that are important for their oversight and grant management functions and the objectives that they need to track to be able to report program effectiveness and impacts.

Assigning the Evaluation Priorities

Once the programs have been sorted into their evaluation groups and the evaluation objectives that SEP managers would like to consider are identified, the evaluation groups, the programs within the groups, and the evaluation objectives must be prioritized. As with most portfolios of programs, the evaluation effort is budgeted consistent with information need and the available resources. As noted earlier rigorously evaluating all of the 400 plus programs would consume the entire national SEP program budget. As a result, only the programs and research objectives that are high priority can be incorporated into the more reliable and costly evaluation approaches. The other programs and objectives must be moved into a less costly and less reliable evaluation approach, or not evaluated at all. This step sets the evaluation priorities that feed into the four different rigor levels associated with the four different evaluation approaches (see next section of this paper).

The assignment of evaluation priorities is always a subjective process and relies on the combination of skills found in the evaluation professionals and on the needs of the SEP program managers. In this effort, the evaluation contractor meets with the SEP program managers and moves down the evaluation groups and assigns evaluation priorities to the all of the evaluation groups. This is done before the priorities are assigned to the programs in the groups or to the

researchable issues. The result of this effort is a prioritized list of evaluation groups. This process considers the type of programs in the group, the size of the group, the size of programs in the group, the total program funding level of the group, the need for energy impact, and other data for that group.

Next, the programs within each group are prioritized in much the same way. Each group is examined separately, so that the programs are prioritized within each group using much of the same information that was used to prioritize the groups. The result from this effort is a prioritized list of programs within each evaluation group.

Lastly, the researchable issues within each program group are prioritized based on the information needs of the SEP managers and the oversight and policy issues with which the managers face.

Typically, a priority ranking approach uses a numeric scale, such as a 1-to-5 scale, or a 1-to-10 scale where 10 represents a very high priority and a zero or 1 represents a very low priority. Priorities can also be assigned as very high, high, medium, low and very low. The result of this effort is a priority list of research groups, programs and researchable issues on which the studies can focus. However, once the priority lists are established, they should be considered guidelines rather than prescriptive assignments. It is possible that a program can have a low group score, a medium program score and a very high research objective score. As noted in the review by Deloitte, there may be times when little information is known about a small program that may have a high national potential. In these cases, the very high scoring objective may be great enough to incorporate that program evaluation objective into a higher rigor approach. In fact, one of the criteria for assigning priority scores can be the potential for the program to have a high national impact even though the current program may be small.

Conducting the Evaluations – A Four Priority Approach

Once the evaluation priorities are established they can be sorted into various evaluation rigor approaches. In this paper, we recommend the use of four different evaluation priorities in which an evaluation approach is assigned to three of the priority groupings. An overview graphic of this approach is provided in Figure 1.

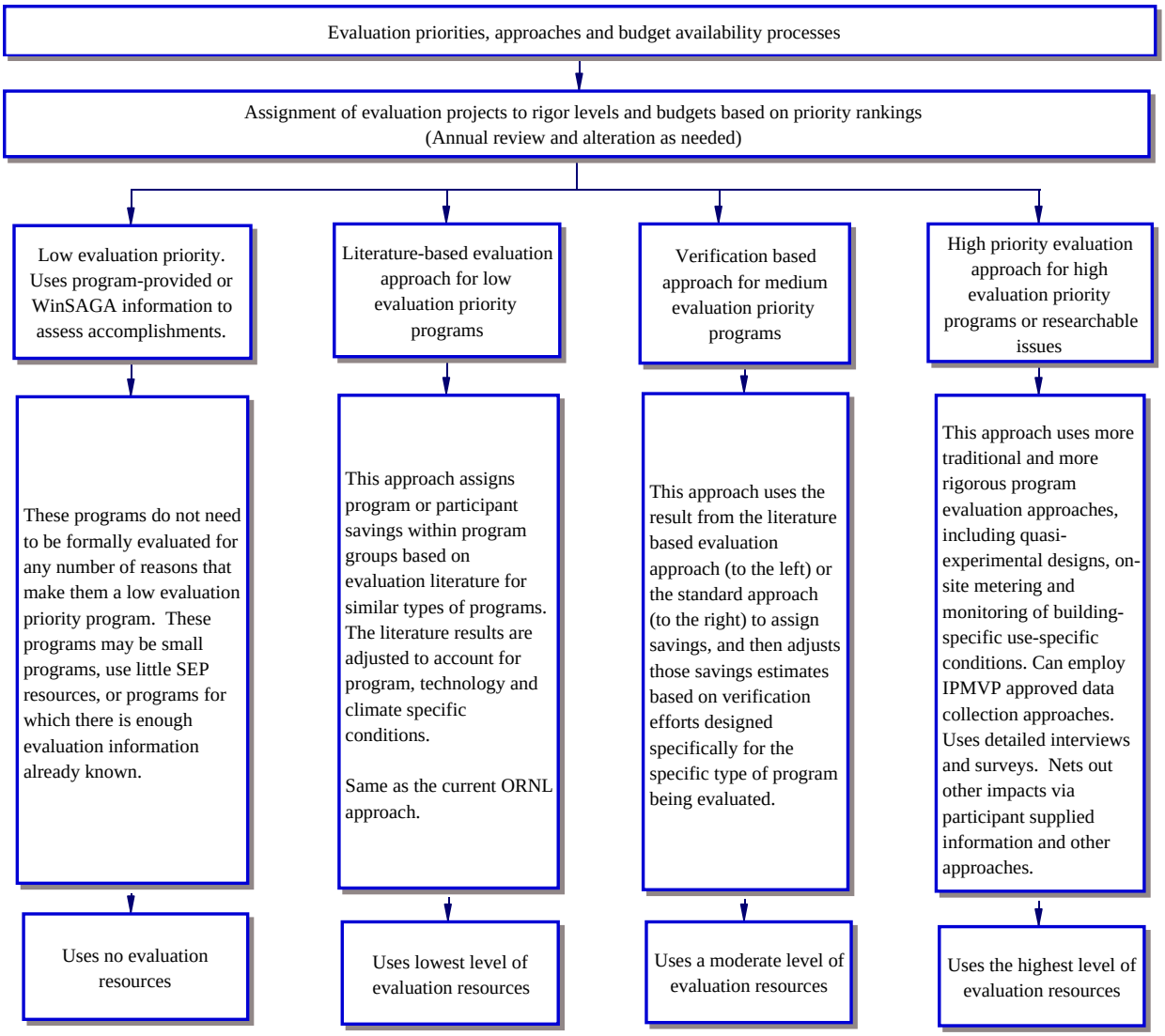

Figure 1 SEP program evaluation approach and priority levels

The evaluation approach decision is based on the reliability of the information needed and the priority of the evaluation group, the individual program and the researchable issues. These are discussed in more detail below and include a low evaluation priority group for the lowest priority programs that SEP managers do not need to evaluate for various reasons. This may because the programs are too few to represent a substantially contributing part of the SEP portfolio, may be programs that are already well understood and for which energy impact data is easily available and is considered accurate, may be programs that are not expected to provide significant energy savings or are considered low priority programs for other reasons.

The non-evaluation group is followed by a low-rigor priority group for programs that need to be evaluated, but do not need to be assigned higher cost approaches more appropriate for higher priority programs, but are important enough to be assigned some level of evaluation effort. This group of programs can be evaluated using a low-cost literature-based evaluation approach similar to the approach used by ORNL in their June 2005 study. Energy savings are based on more

rigorous studies conducted of SEP programs within the higher priority studies or on studies of other program evaluations found in the literature with savings estimates adjusted for SEP-specific program operations and conditions identified during the early evaluation planning efforts and the associated program-specific data collection efforts. This approach is similar to the ORNL approach, but is based on a more complete understanding of the SEP programs and their operational conditions so that program-specific adjustments to the savings found in the literature can be applied to the SEP programs included in this group. Under this approach the savings adjustments are based on the weather-specific, technology-specific conditions and key program operational differences between the SEP program being evaluated and the literature-based study findings.

The next group is the medium-rigor approach, called the verification approach. In this approach, verification surveys, interviews or on-site visits are performed to confirm the changes caused by the program and to adjust energy impacts estimates to reflect the conditions observed or recorded in the verification efforts. As with the previous group, the energy impacts for this group can be literature based or can be based on the results of the high-priority higher-rigor study results for similar types of programs, depending on availability of the evaluation information and the studies that undergo a more rigorous analysis. In this group reliability is enhanced via the verification efforts conducted at the program level which are then linked higher rigor evaluation results found in the literature or conducted via the higher priority evaluation efforts described in the standard evaluation approach priority.

The high priority evaluation group is reserved for the higher priority evaluation needs and is expected to be assigned to programs and program groups that need to be evaluated via a more traditional evaluation approach typically associated with the energy efficiency and renewable energy program evaluation field. The high priority approaches include a wide variety of program-specific, participant-specific, facility-specific evaluation approaches that are designed for a particular high priority SEP program or a group of similar high priority programs. These studies employ experimental or quasi-experimental designs more typically associated with energy programs, as well as collect on-site and off-site participant-based program effects information. These studies also can employ on-site metering, monitoring or performance verification efforts that help reduce uncertainty associated with the evaluation efforts.

In summary, the four rigor-level approach uses the following evaluation approaches:

1. Non-evaluated priority programs. Programs that are very low on the evaluation priority assignment scale. These programs may not need to be independently evaluated, allowing OWIP and SEP managers to rely on program-reported accomplishments or WinSAGA information to identify program activities, successes and accomplishments. For these programs independent evaluations may not be needed.

2. Literature-based program evaluation priority. This priority group is the lowest priority for which evaluation resource are allocated and is used for programs that need to be evaluated, but are not a high enough priority to undergo field verification efforts or require more rigorous and more costly efforts normally associated with more traditional evaluation approaches. This approach assigns energy impacts based on program

evaluation results found in the literature to match the program conditions associated with the SEP programs being evaluated. The primary effort in this group is identifying evaluations presented in the evaluation literature or conducted via the SEP high priority approaches and making adjustments to those energy impacts based on program structure, function, operations, target markets, services provided, weather conditions and other impact adjustment metrics. The end result of this evaluation approach is an estimate of energy impacts using pre-existing evaluation results as the platform from which adjustments to those impacts are applied. This is the same approach used by ORNL in their June 2005 evaluation study.

3. Verification-based program evaluation priority. This medium-rigor approach also uses the literature to assign program impacts similar to the lower-rigor literature based approach, or uses the evaluation results from the higher-rigor standard SEP evaluation approaches of similar programs and then adjusts the results in accordance with the findings from the field verification efforts conducted to confirm the program services and delivery results. In this approach, a number of verification approaches can be used depending on the program and the approach needed. Verification can be conducted using on-site examinations of program operations to confirm service delivery efforts, approaches and services, on-site examinations of participant's facilities to confirm operational conditions and equipment used to acquire the energy impacts, telephone interviews or surveys with participants to confirm information on which the energy impacts are grounded, internet or e-mail data or service confirmation efforts or other approaches. These approaches are described in the California Evaluation Framework and the California Evaluation Protocols. The verification approach is one-step-up from the literature-based approach in that it adjusts the results of program evaluations found in the literature or conducted via the high-rigor SEP evaluations, but trues-up the energy estimates by using the results of the field verification efforts to confirm the results of the programs efforts. The programs in this group are high enough priority that using only literature based energy impact adjusted estimates are not sufficient to meet the evaluation reliability standards for this group of programs.

4. High priority evaluation approach. This is the highest rigor level associated with the SEP program evaluation effort. The high priority approach uses program-specific evaluation efforts that are more traditionally applied within the energy program evaluation community. In this approach, the highest priority programs are evaluated using quasi-experimental designs or designs that are specifically tailored for the individual programs being evaluated. This approach uses on-site inspections, participant and non-participant interviews, interviews with program managers, partners, and stakeholders. These approaches often use regression analysis techniques and test for differences between pre-program and post-program energy use levels adjusted by a reference baseline or comparison group. Metered data or the use of utility metered data is often analyzed for participant impacts. These studies use engineering and building modeling techniques and sometimes employ isolated retrofit measurements or other IPMVP metering and monitoring approaches. These studies typically test for both gross and net energy impacts. Renewable energy programs are tested not only for capacity but also for energy produced from renewable energy facilities

The objective of this four-tiered evaluation priority approach is to spend evaluation dollars appropriate to obtain the reliability needed within the researchable issues to be addressed within the SEP evaluation objectives. Thus, the approaches identified above are linked directly with budget availability. The non-evaluation group needs no evaluation budget and is established in order to assign programs that do not need to be evaluated. The literature-based group needs only the level of budget to research the literature and allocate savings across similar types of SEP programs and make engineering or building modeling adjustments to the savings based on specific program, technology and climate conditions. The verification approach needs a budget that can support applying literature-based findings or the findings from higher-rigor standard approaches, plus the cost of conducting the necessary field verification efforts and adjusting savings to reflect the verification results as well as the results associated with the programs design and operational conditions as well as weather and other adjustments appropriate for the program being evaluated. The highest level rigor is associated with the high priority evaluation approach will employ traditional evaluation planning efforts in which the evaluation is designed to match the program or program group's conditions. These approaches consist of the more traditional quasi-experimental design approaches, metering and monitoring efforts and data analysis approaches typically used in the energy program evaluation industry. These approaches are discussed in detail within such documents as the California Evaluation Protocols and the associated metering, monitoring and verification efforts presented in this document.

This four priority approach allows the majority of the evaluation resources to be placed on traditional evaluation approaches that require the greatest proportion of the research budget, and deliver the needed reliability in the study results, while at the same time allowing for the use of less rigorous evaluation approaches associated with lower evaluation priority programs. This approach is also consistent with the Deloitte recommendations to establish an evaluation approach that provide more reliable evaluation results for *the most important, most costly or least well understood programs,* while at the same time providing flexibility in the evaluation planning process to incorporate researchable issues beyond energy impacts when desired by the SEP program managers and policy oversight staff.

References

TecMarket Works Team, *California Evaluation Framework,* June 2004.

California Public Utilities Commission, TecMarket Works Team, *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals,* April 2006.

Thomas D. Cook and Donald T. Campbell, *Quasi-Experimentation: Design & Analysis Issues for Field Settings,* 1979. Houghton Mifflin Company: Boston.

Related Publications

AIS, SRC International (2001). *European Ex-Post Evaluation Guidebook for DSM and EE Services Programmes*. International Energy Agency. April.

ASHRAE (1999). Chapter 39: Building Energy Monitoring. *ASHRAE Handbook of Fundamentals*. ASHRAE: Atlanta, GA.

ASHRAE (2001a). *ASHRAE Standard 14-2001 - Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs*. ASHRAE.

ASHRAE (2001b). Cooling Load Calculations (Chapters 28 and 29). *ASHRAE Handbook of Fundamentals*. ASHRAE: Atlanta, GA. I-P Edition.

ASHRAE (2001c). *Method of Test for Determining Design and Seasonal Efficiency of Residential Thermal Distribution Systems, Third Public Review Draft*. American Society of Heating, Refrigeration and Air Conditioning Engineers. ASHRAE Standard 152-P.

ASHRAE (2002). Measurement of Energy and Demand Savings, Guideline 14. American Society of Heating, Refrigeration and Air Conditioning Engineers: Atlanta, GA.

Bronfman, Benson and Jane Peters (1991). *Process Evaluations of DSM Programs*. Oak Ridge National Laboratory. December.

CADMAC (1998). *Quality Assurance Guidelines for Statistical, Engineering, and Self-Report Methods for Estimating DSM Program Impacts*. Appendix J.

CADMAC (1999). *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs*. California Public Utilities Commission. 93-05-063.

California State Governor's Office (2001). *Standard Practice Manual (SPM): Economic Analysis of Demand-Side Management Programs*. October.

Campbell, Donald and Julian Stanley (1963). *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin: Boston, MA.

Conant, Dorothy and Jeremy Schutte (1993). "Matching Methodologies and Program Types: An Evaluation Retrospective." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 155-160.

Davidson, Jane E. (2000). "Ascertaining Causality in Theory-Based Evaluation." *New Directions for Evaluation* (Number 87): 17-26.

DeCotis, Paul A, Mark C Coleman, Jennifer Ellefsen and Helen Kim (2000). "Portfolio Approach to Designing and Evaluating Buildings Energy Efficiency Programs." *National Energy Services Conference*: Lake Worth, FL. pp. 495-503.

DeCotis, Paul A. and John F. Munro (2001). "A Theory-Based Systems Approach for Evaluating Energy Efficiency Investments Portfolios." *National Energy Services Conference*: Lake Worth, FL. pp. 282-296.

DeCotis, Paul A, Larry Pakenas, and Joel Eisenberg (2002). "Systems-Based Portfolio Evaluation: Diagnostic Benefits and Methodological Challenges." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 10.57-10.67.

EPRI (1992). *DSM Process Evaluation: A Guidebook To Current Practice*. Electric Power Research Institute. Project RP2981-2.

Erickson, Jeff and Oscar Bloch (2002). "Using Program Theories to Align Performance Metrics with Public Purpose Goals." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 5.65-5.75.

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.

Goldberg, Miriam and Edmund Kademan (1995). "Is it Net or Not? A Simulation Study of Two Methods." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 459-465.

Groves, Robert M, Floyd J Fowler, Mick P Couper, James Lepkowski, Eleanor Singer and Roger Tourangeau (2004). *Survey Methodology*. Wiley: Hoboken, NJ.

Hall, Nick and Gretchen Jordan (2001). *The 2001 FEMP Customer Survey: Study Report*. Sandia National Laboratories.

Hall, Nick and John Reed (1997a). "Methods for Measuring Customer Satisfaction." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 23-33.

Hall, Nick and John Reed (1999). "Market Transformation: Expectations vs. Reality." *Home Energy Magazine*. July 1999. 16-20.

Hall, Nick and John Reed (2001). "Merging Program Theory and Market Theory in the Evaluation Planning Process." *International Energy Program Evaluation Conference*: Salt Lake City, UT. pp. 297-304.

Hastie, Steve, Ralph Prahl, Phil Mosenthal, Dimple Gandhi and Barbara Klein (2000). "A Systematic Application of Theory-Based Implementation and Evaluation of Market Transformation Programs." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 6.157-6.168.

Heckmann, James J (1979). "Sample Selection Bias as a Specification Error." *Econometrica* 47: 153-162.

Herman, Patricia, Shel Feldman, Shahanna Samiullah and Kirsten Stacey Mounzih (1997). "Measuring Market Transformation: First You Need a Story ..." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 319-326.

Heschong Mahone Group (2002). *Time Dependent Valuation (TDV) Economics Methodology*. Pacific Gas & Electric.

Hirst, Eric and John Reed (1991). *Handbook of Evaluation of Utility DSM Programs*. Oak Ridge National Laboratory. ORNL/CON-336.

Nexant and Lawrence Berkeley National Laboratory (2002). *Detailed Guidelines for FEMP M&V Option A*. Federal Energy Management Program.

PA Consulting Group Inc. (2003). *Standardized Methods for Free Ridership and Spillover Evaluation - Task 5 Final Report (Revised)*. National Grid, NSTAR Electric, Northeast Utilities, Unitil, Cape Light Compact. June 16.

Peters, Jane, Bruce Mast, Patrice Ignelzi and Lori Megdal (1998). "Measuring Market Transformation: The 1997/1998 California Market Effects Studies." *National Energy Services Conference*: Orlando, FL. pp. 121-128.

Phillips, Jack J (1997). *Handbook of Training Evaluation and Measurement Methods*. Butterworth-Heinemann: Woburn, MA.

Phillips, Jack J and Ron Drew Stone (2002). *How to Measure Training Results: A Practical Guide to Tracking the Six Key Indicators*. McGraw-Hill: New York, NY.

Pigg, Scott and Michael Blasnik (1993). "Dealing with Outliers in Impact Evaluations Based on Billing Data." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 188-198.

Quantum Consulting Inc. (2004). *National Best Practices Study*. California Measurement Advisory Council (CALMAC).

Rasmussin, Tami, Kathleen Gaffney and Rob Rubin (2003). "Addressing Program Attribution in the Wake of the California Energy Crisis." *International Energy Program Evaluation Conference*: Seattle, WA. pp. 443-453.

Reed, John, Mary O'Drain and Jim Chace (1999). "Transforming Markets Through Education and Information: A Study of the Pacific Energy Center." *International Energy Program Evaluation Conference*: Denver, CO. pp. 841-855.

Reed, John, Innovologie, Gretchen Jordan, Sandia National Laboratory, Ed Vine, Lawrence Berkeley National Laboratory, *Impact Evaluation Framework for Deployment Programs: An Approach for Quantifying Retrospective Energy Savings and Market Effects*, July 2007.

Ridge, Richard (1997). "Errors in Variables: A Close Encounter of the Third Kind." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 479-487.

Ridge, Richard, Kathleen McElroy and Rob Rubin (2001). "Testing the Causal Linkage Between Training of Sales Personnel in Retail Lighting and Appliance Stores and Changes in Market Share of ENERGY STAR® Qualifying Equipment." *International Energy Program Evaluation Conference*: Salt Lake City, UT. pp. 251-261.

Ridge, Richard, Daniel Violette and Donald R. Dohrmann (1994). *An Evaluation of Statistical and Engineering Models for Estimating Gross Energy Impacts*. California Demand Side Management Advisory Committee: The Subcommittee on Modeling Standards for End Use Consumption and Load Impact Models. June.

Ridge, Richard, Daniel Violette and Donald R. Dohrmann (1998). Appendix J: Quality Assurance Guidelines for Statistical and Engineering and Self-Report Methods for Estimating DSM Program Impacts. *CADMAC Evaluation Protocols*. CADMAC.

Rogers, Patricia, Timothy A Hacsi, Anthony Petrosino and Tracy A Huebner (2000a). "Program Theory in Evaluation: Causal Models in Program Theory Evaluation." *New Directions for Evaluation* Fall (87): 47-55.

Rogers, Patricia, Timothy A Hacsi, Anthony Petrosino and Tracy A Huebner (2000b). "Program Theory in Evaluation: Challenges and Opportunities." *New Directions for Evaluation* Fall (87).

Rogers, Patricia, Timothy A Hacsi, Anthony Petrosino and Tracy A Huebner (2000c). "Program Theory in Evaluation: Practice, Promise, and Problems." *New Directions for Evaluation*.

Rossi, Peter, Howard Freeman and Mark Lipsey (1999). *Program Evaluation: A Systematic Approach*. 6th Edition. Sage Publications: Thousand Oaks, CA.

Rubin, Herbert J (1983). *Applied Social Research*. Bell & Howell Company: Columbus, OH.

Rufo, Mike (1993). "DSM Resource Planning the Next Generation: Building the Foundation Through Evaluation." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 162-168.

Sanderson, Ian (2002). "Evaluation, Policy Learning and Evidence-Based Policy Making." *Public Administration* Vol 80 (No. 1): 1-22.

Schiffman, Dean A and Robert F Engle (1993). *Appendix Z Simulation Study: Comparison of Alternative Methods for Measuring the Gross and Net Energy Impacts of Demand-Side Management Programs (with Addendum).* San Diego Gas & Electric.

Schutte, Jeremy and Daniel Violette (1994). "The Treatment of Outliers and Influential Observations in Regression-Based Impact Evaluation." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 8.171-8.176.

Sebold, Frederick D, Alan Fields, Lisa Skumatz, Shel Feldman, Miriam Goldberg, Kenneth Keating and Jane Peters (2001). *A Framework for Planning and Assessing Publicly Funded Energy Efficiency*. Pacific Gas & Electric. March 1. Study PG&E-SW040.

Smith, Peter R, Paul A DeCotis and Karl S Michael (1998). "Linking Market-Based Energy Efficiency Programs to Economic Growth, Sustainable Development and Climate Change Objectives." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 9.185-9.195.

Sonnenblick, Richard and Joe Eto (1995). "Calculating the Uncertainty in Estimates of DSM Program Cost-Effectiveness." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 759-768.

Spellman, Richard, Bruce Johnson, Lori Megdal and Shel Feldman (2000). "Measuring Market Transformation Progress & the Binomial Test: Recent Experience at Boston Gas Company." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 6.367-6.378.

Sumi, David, Jeff Erickson and Jim Mapp (2002a). "Wisconsin's Public Benefits Approach to Quantifying Environmental Benefits: Creating Different Emissions Factors for Peak/Off-Peak Energy Savings." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 9.353-9.363.

Sumi, David, Bryan Ward, Glen Weisbord and Mike Sherman (2002b). "Quantifying Economic and Environmental Benefits." *National Energy Services Conference*: Jupiter, FL. pp. 7.149-7.163.

TecMRKT Works (2001). California Low Income Public Purpose Test.

Tolkin, Betty and Glenn Reed (1993). "Free-Ridership Estimation in the New Construction DSM Market." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 787-791.

Train, Kenneth (1994). "Estimation of Net Savings from Energy Conservation Programs." *Energy* Vol 19 (No. 4): 423-441.

Train, Kenneth, Susan Buller, Bruce Mast, Kirtida Parikh and Eric Paquette (1994). "Estimation of Net Savings for Rebate Programs: A Three-Option Nested Logit Approach." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 7.239-7.247.

USEPA (1995a). *Conservation and Verification Protocols, Version 2.0*. U.S. Environmental Protection Agency. 403/B-95-012.

USEPA (1995b). *Conservation Verification Protocols: A Guidance Document for Electric Utilities Affected by the Acid Rain Program*.

Vine, Ed and Jayant Sathaye (1999). *Guidelines for the Monitoring, Evaluation, Reporting, Verification and Certification of Energy-Efficient Projects for Climate Change Mitigation*. Lawrence Berkeley National Laboratory. March.

Vine, Ed and Jayant Sathaye (2000). The Monitoring, Evaluation, Reporting, Verification, and Certification of Energy-Efficiency Projects. *Mitigation and Adaptation Strategies for Global Change*. Kluwer Publishing: Dordrecht, Netherlands. 189-216.

Violette, Daniel (1991). "Impact Evaluation Accuracy and the Incorporation of Prior Information." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 86-92.

Violette, Daniel (1995). *Evaluation, Verification, and Performance Measurement of Energy Efficiency Programs*. International Energy Agency.

Violette, Daniel, Richard Brakken, Andy Schon and Jerry Green (1993). "Statistically Adjusted Engineering Estimates: What Can the Evaluation Analyst Do About the Engineering Side of the Analysis?" *International Energy Program Evaluation Conference*: Chicago, IL. pp. 656-663.

Weiss, Carol H (1998). *Evaluation: Methods for Studying Programs and Policies*. Prentice Hall: Upper Saddle River, NJ.

Weiss, Carol H (2000). "Which Links in Which Theories Shall We Evaluate?" *Program Theory in Evaluation* Fall (No 87).

Woods, James and Sami Khawaja (2000). "Multi-Attribute Valuation for Cost Effective Evaluation of Market Transformation and Other Hard to Quantify Programs." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 4.227-4.238.

Worrell, Ernst, John Laitner, Michael Ruth and Hodayah Finman (2001). *Productivity Benefits of Industrial Energy Efficiency Measures*. Lawrence Berkeley National Laboratory. December.

Worthen, Blaine R, James R Sanders and Jody L Fitzpatrick (1997). *Program Evaluation: Alternative Approaches and Practical Guidelines*. 2nd Edition. Longman: New York, NY.

Wright, Roger (1983). "Finite Population Sampling with Multivariate Auxiliary Information." *Journal of the American Statistical Association* Vol. 78: 879-884.

Wright, Roger and David Jacobson (1993). "A Methodology for Integration of Evaluation Studies." *National Demand-Side Management Conference*: Miami Beach, FL. pp.

Xenergy (1996). *Net Savings Estimation: An Analysis of Regression and Discrete Choice Approaches*. CADMAC Subcommittee on Base Efficiency.