

Speaking the Same Language: Promoting Evaluation-Readiness in Program Design and Implementation

Crispin Wong and Kerstin Rock, PECE

ABSTRACT

Program implementation and evaluation—by design—take on very different objectives within the demand side management (DSM) program cycle. Program implementation focuses on developing and bringing to market effective and innovative solutions capable of delivering energy efficiency and/or demand response targets. Evaluation, on the other hand, is tasked with estimating program impacts and supporting program improvement. Even though the two parties have coexisted in the industry for a long time, there are still barriers and constraints impeding the effective and timely transfer of knowledge regarding timing, access to data, and regulatory restrictions on communication. As the industry matures and faces more complex implementation and evaluation, it is crucial to understand why effective integration of evaluation results into program design and implementation continues to remain such a challenge and what we (evaluators and implementers) can do to fix it. We have combined our perspectives as evaluation contractors and internal EM&V staff at an implementation firm to provide a unique look at where disconnects between evaluators and implementers typically happen and why they occur. We also offer recommendations regarding basic steps both sides can take to speak a common language and achieve mutual goals.

Introduction

The difficulty in performing evaluations and closing the loop between evaluators and implementers is a well-known and long discussed issue within the demand side management (DSM) program community. Evaluation has historically taken place at the end of a program cycle, after implementers and utilities have moved onto next generation programs. This time lag makes it particularly difficult to excavate the right information and if it does not exist, to recreate the record. Evaluators have suggested early evaluation or evaluability assessments as solutions to better streamline implementation and evaluation, but even these approaches have barriers (Bronfman et al., 2008). The industry has matured to a point where both evaluation and implementation are increasingly complex. Programs have been in markets longer; easy to implement measures have saturated markets; targets are increasing; and the economic climate necessitates that programs achieve more with less funding. Programs are using increasingly layered program approaches with a variety of measures and O&M activities to find more savings, while the number of funding sources and market actors continue to expand and complicate savings attribution (Skumatz et al., 2009).

As a result, evaluation has become more complex and must address more issues related to attribution and net savings with smaller budgets (Skumatz et al., 2009). Many evaluators have correctly argued that they would be more able to effectively tackle complex evaluation issues if they had access to proper documentation and correct and accurate data. Likewise, implementers and administrators have stated that they would be more willing and able to address EM&V concerns if program teams better understood the evaluation approaches applied to programs and

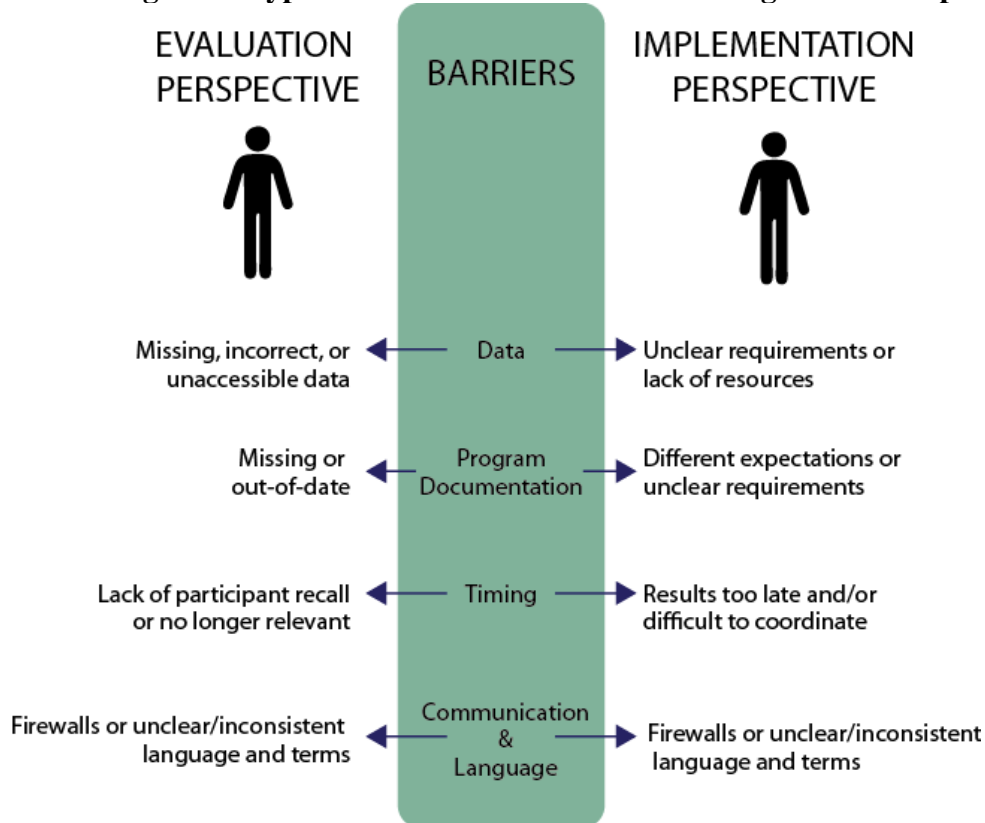
could reasonably anticipate the data and documentation requirements during program design and/or early program launch.

This paper draws on the authors' experiences as third-party evaluation contractors and internal EM&V support at PECEI, an implementation contracting firm. This combination provides us with a unique view of evaluation issues from both evaluator and implementer perspectives. This paper is focused on identifying common challenges faced by both sides and provides simple (but not necessarily easy) steps both camps can take to improve program efficacy and reduce all too common administrative inefficiencies. We first outline typical challenges that have led to both parties feeling generally misunderstood and underappreciated by the other. We then review past efforts and existing industry solutions to address these issues. Last, we provide recommendations and solutions for both camps--based on our firm's efforts to integrate EM&V into program design and implementation as well as our efforts to extract more value from EM&V studies. In short, this paper offers a practical perspective on how to help our industry overcome the disconnects between implementation and evaluation processes.

The Problem

Common evaluation issues and disconnects from the evaluator and implementer perspectives are shown in Figure 1. These include data, program documentation, timing, and communication and language.

Figure 1. Typical Barriers To Effective Working Relationships



Data availability and quality is essential for evaluators to contact the right participant representative, verify program measures, and confirm program performance. A long standing issue is that implementers do not have the same data collection objectives as evaluators and may not collect data necessary for evaluation (West & Bronfman, 2009). Implementers, on the other hand, may not understand all the data requirements, have time to collect data at the correct detail or frequency or be able to anticipate or understand evaluation needs. Evaluation requirements or advance information on evaluation approaches is not readily available, so even if implementers try to prepare for evaluation, there is no guarantee they will capture the correct data requirements. When this information does become available, it is usually past the time when the program team can adequately address any shortcomings. Implementation records, conversely, may not be clear or easy for evaluators to reconcile since they were not involved in implementing projects. While application of professional judgment will always be necessary to some extent, we believe that the frequency with which it is applied could be greatly reduced if better data is available to evaluators.

Documentation of the program framework (e.g. logic model¹), or lack thereof, also frequently poses a barrier. Evaluators often extract these details through staff interviews and then compile the information to develop an evaluation framework. This can lead to knowledge gaps or misunderstandings. In addition, the people who design programs are not always the same as those who implement, and program design documentation does not always reflect how the program was actually implemented.

Communication breakdowns between evaluators and implementers can also be a barrier. In some cases this is because they are not allowed to talk due to firewalls. Other times, evaluators and implementers use a different vocabulary and reference point, which leads to misunderstandings. Implementers may misunderstand what evaluators are asking for or are unsure how much information they can hand over to evaluators while still maintaining proper distance. In addition, evaluators often fail to clearly discuss what the evaluation results will be used for or how they can provide more value than an “audit” function. This can lead to a tense relationship based on fear and frustration, rather than one that is educational and collaborative.

Closing the loop between implementation and evaluation has been the other significant hurdle in this industry (Collins & Bishop, 2009). Evaluations attempt to review what happened in order to reconcile previous program results and hopefully inform future programs. In many cases, the evaluation occurs too long after implementation to fully address either purpose. Evaluation reports may not provide recommendations that can be realistically implemented or ex post values that can facilitate future program improvements.

So where and why do these issues arise? The disconnect between evaluators and implementers likely stems from three key sources. First, evaluators and implementers have different job objectives by design. Evaluators, to some degree, represent the rate payer perspective and ensure the resources claimed are real and cost-effective. They have an incentive to ensure savings are not overstated and that programs have spent funds wisely. Implementers are often paid based on achieved savings and thus have an incentive to maximize the savings credited to the program. The competitive environment also dictates that programs are run efficiently, are innovative, and employ continuous learning processes.

Second, the two parties often come from different backgrounds that shape how they think about programs, data collection and analysis and applicable success indicators. In our observation, evaluators tend to come from an academic background focused on economics,

¹ For more information on logic models, refer to <https://www.bja.gov/evaluation/links/WK-Kellogg-Foundation.pdf>

mathematics or other technical background. Implementers frequently are more operational and industry based, with a focus on business or other related topics. Engineers are common to both groups but are influenced by the aforementioned job objective differences. While these are simple anecdotal observations, they help explain why communication barriers exist. Let's be honest—while some improvements have been made, many evaluators still turn out lengthy reports that are difficult for anyone besides the authors to read. While it is necessary to document in detail the assumptions, methods and data used to develop the evaluation, many reports lack an accessible summary written for audiences without in-depth knowledge of statistics, sampling, or other verification approaches. The complexity of reports coupled with everyone's time constraints frequently results in limited readership of draft or final reports.

Third, to eliminate real and perceived conflicts of interest, evaluators and implementers are often kept separate in jurisdictions enforcing strict firewalls between the two parties. The inability to coordinate and communicate throughout the program cycle impedes evaluation and timely feedback needed for adaptive management (Messenger et al. 2010). Evaluators face barriers in getting the right data, starting early enough to provide timely feedback and gaining access to information and projects to accurately assess what happened. Implementers are confused when there is little direct communication from evaluators and sometimes do not understand the messages even when they do communicate.

Finally, another important consideration is the growth of the industry and number of people new to utility and other DSM programs. Many new hires filling utility evaluation, program administration, third-party implementation, and evaluation positions are new to the industry, frequently lacking the necessary background and training to quickly get up to speed. While the industry has recognized the general need for industry and technical training and is taking steps to address these needs, is it not uncommon to have one or more individuals with minimal expertise and/or knowledge of the utility industry (economics, regulation, program implementation or EM&V) involved in a particular program evaluation effort (either on the side of the utility, implementer or EM&V contractors). Combined with the other barriers discussed, this constitutes a significant institutional challenge and if left unaddressed, will continue to hamper the industry's ability to meet ever increasing goals with lower budgets.

Existing Solutions

Evaluators have promoted early evaluation or evaluability assessments to prepare program implementers for evaluation and, ideally, catch any problems early in the implementation process. Specifically, this approach determines what data should be tracked, documents the program framework, and reviews initial implementation processes to ensure an evaluation will be successful (Bronfman et al., 2008). Past studies on evaluability assessments found that two key things need to happen in order to realize their full value. First, the implementer needs to understand the issues identified by the evaluator and how it relates to evaluation. Second, the implementers must follow through on the evaluability assessment recommendations (West and Bronfman, 2009). This is easy to say, but much harder to implement if there is no close relationship between the evaluator and implementer. Arguably the two parties will rarely achieve this level of communication because of conflict of interest requirements and natural separation.

It is also important to consider that program implementation moves at a fast pace. In order to have traction with a program, the recommendations or action items must be easy to

understand and implement, and must generally add value without increasing costs. Fleeting interactions with a third-party evaluator is not enough time for implementers to absorb the context and action items in a meaningful manner.

Early evaluation or evaluation concurrent to the program cycle is helpful in promoting timely feedback. States such as Michigan and Pennsylvania implement this framework as well as organizations like Bonneville Power Administration (Cadmus Group, 2010) and the Northwest Energy Efficiency Alliance. However, the parameters of successful evaluability assessment still apply here. Implementers must understand the context and know what to do with the information, and evaluators must provide suggestions that can be implemented in a timely manner. In addition, program administrators and utilities must plan for concurrent evaluation in budgets and staff availability, and in some cases, gain regulatory approval to do this. Concurrent evaluation can be logistically complex, with constant information transfers between evaluator and implementer, and time lags still exist because most jurisdictions complete evaluations after a program year or cycle comes to a close.

New Solutions: Closing the Gap

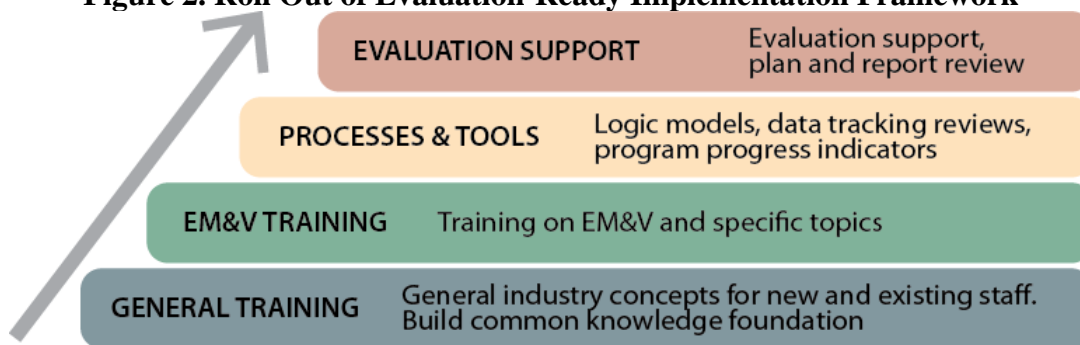
The barriers, industry context, and slow adoption of existing solutions, such as evaluability assessments and early evaluation, result in industry inefficiencies. Leveraging our experience as evaluators and implementers, the following section highlights several concrete recommendations for addressing these challenges. We first review activities to incorporate into implementation and/or program administration, followed by suggestions for evaluators. Many of the activities discussed are low-cost, high benefit activities that the industry can put into practice.

Solutions for Administrators and Implementers

In the absence of timely and sufficiently relevant EM&V input or data from third party evaluators, program administrators and implementers should consider developing evaluation-ready program frameworks. Actively including these activities in program implementation can save time and effort during evaluation for both the implementer and the evaluator. Figure 2 presents a high-level overview of a process for integrating evaluation concepts into day-to-day implementation activities. PECI has implemented and refined this process over the past two years, with successes, challenges, and lessons learned along the way.

We found that effective solutions start with solid industry knowledge as well as an understanding of EM&V and implementation-specific concepts and terminology. Based on this foundation, processes and tools can be developed and deployed. Finally, we provide assistance and mentoring throughout the evaluation process. While we provide examples from our experiences at a particular third-party implementation firm, we believe our observations apply to both the program administrator and implementer and are transferable to other situations.

Figure 2. Roll Out of Evaluation-Ready Implementation Framework



Training

Training is by no means a new idea, but it remains a key component to promoting evaluation efficiency. It is even more important in the context of industry growth that necessitates recruiting EM&V and program staff from other industries into utilities, implementation firms, and evaluation consultancies. The diversity of backgrounds and industry-specific knowledge requires creating a common energy industry knowledge foundation that covers industry fundamentals. A detailed review of available industry trainings in 2009/2010 revealed that while training opportunities are plenty, most focus on specific applications and/or target audiences. None provided the combination of topics and appropriate depth required to build the necessary foundation we seek. As a result, PEGI developed a training curriculum covering topics including: industry history, utility economics and business models, regulatory processes, basic electricity and gas concepts as well as an overview of energy efficiency EM&V. We also developed and rolled out more in-depth trainings on specific EM&V topics such as impact and process evaluations, specific evaluation approaches (e.g. behavioral programs), logic models, and statistics and sampling.

PECI found that training senior management, business development and program teams resulted in better awareness and familiarity with EM&V and its industry role. For example, after attending an EM&V training, a program team approached us (the internal EM&V staff) to help redesign their program application forms to ensure that the fields captured critical evaluation data about participants. These efforts that arise from a better understanding of evaluation improve program efficacy and future evaluation outcomes.

Processes and Tools

Being evaluation-ready is a shared responsibility within PEGI. Bringing about organizational change and gaining wide-spread staff support requires useful processes and tools. As such, we focused on institutionalizing logic models and data tracking reviews. While it might seem obvious that the people implementing programs should create logic models, many evaluators know that in reality this rarely happens. Evaluators use the logic model as a tool to help identify key performance indicators and an appropriate process evaluation plan. One of the first tasks in many evaluation contracts is to create a logic model, typically based on staff interviews and document review. For a number of reasons, including elapsed time, the skill of the interviewer or the interviewee, or staff transitions, logic models created in this fashion are

frequently inaccurate. Moreover, without any formal training or knowledge of logic models it is difficult for implementers to extract value or even understand how and why evaluators use them as tools. We have found that it is not uncommon for logic models to be mistaken for process flow diagrams, resulting in inaccurate reading and interpretation of the content.

Based on our initial conversations with program staff, it was clear that the key barrier to engaging implementation teams in developing logic models was a lack of understanding of their value to program staff. We found that logic models serve three primary purposes for program design and implementation. First, it supports program design and provides a clear picture of how specific program activities lead to the targeted outcomes. Second, completing logic models and a corresponding indicators table provides program teams with important, relevant data to track in addition to the usual program indicators—kWh, kW, therms and number of measures. Third, it provides a systematic approach for teams to discuss the program framework and come to an agreement on how the program should be implemented to achieve its goals. In one case, when we worked with a program team to document the program design in a logic model, we found that a few design ideas were missing logical links to the outcomes. As a result, the implementation team was able to adjust and augment the program activities to better achieve the outcomes.

The second process addressed was data tracking. Data tracking, as noted, is a crucial part of both monitoring implementation progress and evaluating program success. In an implementation environment, much of the data specification occurs upfront but edits occur to the data structure throughout the program as measures are added or subtracted, new markets are targeted, or measure requirements change. An initial review during an evaluability assessment is useful, but it assumes that the program team knows how to fix any issues, and that they can maintain this on their own as the program requirements change over time. Data tracking has proved to be an unwieldy process, but in many cases, we have been able to apply both incremental and significant improvements to our data collection systems. We recognize that room for improvement remains, and view this as a continuous improvement process.

We found that providing data tracking support on an ongoing basis, where internal evaluation staff is available to answer questions as they arise, is a more efficient way of addressing these issues. Bringing this support in-house (or closer to program implementation) is an alternative approach to hiring an outside company, and avoids the potential disadvantages of a conflict of interest, a cost barrier, or lack of trust.

Evaluation Support

Evaluation is not consistent across programs and regions and can apply different approaches for the same measure. As such, providing ongoing support to implementers as questions and issues arise is important in preparing programs for evaluation. This allows implementers to ask questions about the process and become more informed on the material they are reviewing. For example, in one situation a program team had questions about the validity of a specific net-to-gross (NTG) approach. We discussed the approaches typically used and determined the evaluation applied a fairly standardized approach for the program measures.

In some cases, these tasks are akin to translating from “evaluator speak” into “implementer speak.” For example, one evaluation proposed end use metering with calibrated simulation. The program team did not fully understand the implications of this and we provided a step-by-step approach description (in layman terms) to the team. This discussion went a step

further to identify how these activities impact implementation (e.g. customer will need to agree to a delay in project implementation to accommodate pre-project metering).

Closing the Loop: Acting on the Findings

The next step, integrating evaluation recommendations, is still the holy grail of evaluation. It is easy to identify the straight-forward corrections, such as an incorrect unit energy savings or a mistake on a piece of marketing collateral. Understanding and applying other implementation recommendations can get tricky. Simply understanding evaluation content goes a long way to closing the feedback loop, but there are other issues that need to be addressed such as the feasibility of evaluator suggestions, changing significant program characteristics mid-stream, or implementing cost adding activities.

Recommendations for Evaluators

Thus far, we have identified the actions implementers and program administrators can take. After working closely with implementers, we have noted a few things we would do better as evaluators. While these suggestions do not apply to all evaluations and evaluation reports, we have found that more often than not, they are relevant. The common theme throughout all of our suggestions is that strict adherence to EM&V protocols or best practices in data collection and analysis will go a long way to ensuring quality results, yet, does very little to ensure the findings are resented in an accessible and useful manner. We believe that specific areas of improvement for evaluator communications may include:

1. Know your audience (better) and tailor messages accordingly;
2. Write user-friendly reports (brief, accessible, explicit);
3. Talk to the implementers; and
4. Make recommendations that are specific to the program and detailed enough to feasibly act upon.

Know Your Audience (Better)

Evaluation reports should take different forms depending on the primary audience – in contrast to the all too common practice of developing one-size fits all EM&V study reports. When the evaluation team is trying to meet deadlines, gather results and identify trends, it is easy to forget who might be reading the evaluation report. Tailoring the language and format can go a long way to ensure everyone is on the same page, from regulators, other evaluators, utility staff, program administrators, and program staff. In addition, it is important to consider how long the program has been in the market. If this is the first or second year of the program, the evaluation must provide more background and basic information than a report for programs that have been around for eight years. Utility evaluation staff can also consider this and guide evaluation teams on deliverable requirements.

Improve Reporting and Communication

Evaluations can provide a wealth of information, but it has little value if it is not shared in a way that is readily accessible. Few people enjoy reading really long, 100+ page reports. Nearly all evaluations have key findings and learnings that can be summarized in an executive summary and backed with a 20-30 page report, which is a more manageable length.

Language in evaluation reports is typically very technical, academic and hard to understand for someone who does not live and breathe evaluation and statistics on a daily basis. It is important to be transparent and write in a way that is accessible to a variety of audiences. Providing simple explanations and communicating technical concepts for a non-technical audience is key – appendices can accommodate the minutiae of econometrics and other analyses.

Clarity and explicitness of the approach, findings and recommendations is also very important. At times it is unclear what activities informed which findings and how exactly the evaluator recommends addressing the findings. Where possible, evaluators should be as plain and transparent as possible about the impact and any action the implementer or program administrator should take as a result. Evaluators should not assume the audience can read between the lines of report results.

Talk to the Implementers

Circling back to ask more questions of program administrators and implementers is highly encouraged when not prohibited by regulatory firewalls. While there are concerns about conflict of interest, most implementers really want to understand how to improve upon their programs. If a concept, document, or data are unclear, circling back with the team may provide insight. Engaging early in evaluation plan development is also important, as implementers may provide insight on whether an evaluation approach may or may not work. As evaluators pursue more empirical data, implementers may be the logical and most efficient partner to collect pre-project metering data. This partnership should not be discounted as we continue to pursue more complex programs and evaluation approaches.

Specific Recommendations

Recommendations often come last in the evaluation process but are just as important as quantifying impacts and summarizing program achievements. Recommendations are often general and non-specific, which is not useful to implementers. In addition, recommendations that only add to a program's cost structure are also difficult to implement. In our experience, implementers are open to making changes but only when the recommendations are clearly articulated and sufficiently specific.

Evaluators can form more useful recommendations by considering and/or commenting on whether a particular suggestion is likely to minimize risk, add value, or whether it can be reasonably implemented. Where possible, providing information on the impact of not implementing a recommendation or the positive impact if it is implemented may help provide context for the implementers. Finally, the recommendations must be timely. Any effort to understand the implementer's timeline and when information can really have an impact on the current or future program design increases our efficiency across the industry.

Conclusions

As the industry matures, we are all looking for ways to improve evaluation processes and efficiency while minimizing risk. The framework and suggestions provided in this paper are basic but important to consider so we can move past known barriers and tackle complex issues. Our efforts to promote evaluation-readiness at PEGI have resulted in better program documentation, robust data tracking, continuous improvement in program design and delivery, and more active engagement in the evaluation process. We have also found that evaluation-readiness requires a continuous improvement process—as new evaluation methods emerge, staff change over, and regulation changes—we must adjust the process accordingly and continue to overcome known barriers so we can focus on solving new (and more interesting) problems.

References

- Bronfman, B.; Samiullah, S.; West, A.; Peters, J.; Gandhi, N.; and Cooney, K. 2008. *Integrating Evaluability Assessment into the Program Planning, Implementation and Evaluation Process: Case Studies from Southern California Edison's IDEEA Program Portfolio*. Paper presented at the 2008 Association of Energy Professionals National Conference, Clearwater Beach, Fl. January 29-31.
- Cadmus Group. 2010. BPA Energy Smart Industrial Program: Early Evaluation Report. Bonneville Power Administration. Portland, Ore.: BPA. Retrieved from: http://www.bpa.gov/energy/n/reports/evaluation/pdf/ESI_Early_Evaluation_Report_2009_0618_FINAL.pdf?wwparam=1336020412.
- Collins, M. & Bishop, A. 2009. *Measuring Program Performance from Cradle to Grave*. Paper presented at the 2009 International Energy Program Evaluation Conference, Portland, Ore. August 12-14.
- Messenger, M.; Bharvirkar, R.; Golemboski, B.; Goldman, C.; and Schiller, S. 2010. *Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs*. LBNL-3277E. Berkeley, Ca.: Lawrence Berkeley National Laboratory.
- Skumatz, L.; Khawaja, S.; and Colby, J. 2009. *Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence of Energy Efficiency Behavior*. Berkeley, Ca.: California Institute for Energy and Environment.
- West, A. & Bronfman, B. 2009. *Magic Ingredients in Evaluability Assessments*. Paper presented at the 2009 International Energy Program Evaluation Conference, Portland, Ore. August 12-14.