A New Energy Efficiency Portfolio - Developing an Expanded Evaluation Approach

Douglas Mahone, Heschong Mahone Group Inc., Fair Oaks, CA
Wim Bos, Sacramento Municipal Utility District (SMUD), Sacramento, CA

ABSTRACT

In 2007, SMUD adopted long-term goals for aggressive energy savings, and has developed an expanded portfolio of efficiency programs to meet these goals. With a wider range of savings to target, SMUD is revisiting the question of how savings are measured. Under its traditional approach, savings have has been relatively easy to measure. The emphasis was on discrete measures, with deemed or identifiable savings, which could be counted and "banked". This approach is very defensible, and has worked well, but will not be adequate for measuring the results of the new portfolio of programs.

SMUD's Board will need broader objectives for evaluation, and different criteria for judging program successes and failures, to meet their broader goals. These may lead to different evaluation protocols.

One way to think about this problem is to ask the question, "Where do we want to draw the boundaries of our influence and our savings?" The question relates to gross savings, to free-ridership, to spillover, and to market effects, as well as to where the boundaries of the programs are drawn. The answer will strongly influence how SMUD staff operate their programs, how SMUD measures the success of the new program portfolio, and how SMUD reports its savings to stakeholders.

These issues are hotly debated by evaluators and policymakers. Some contend that the gross savings are what really matter to system planners and procurement needs; and that net savings are only needed for determining shareholder earnings and for improving program implementation. In some jurisdictions, evaluators only measure free-ridership, while not recognizing free-drivers or spillover; there is concern that this is asymmetric and unfairly punitive for program administrators. Some people have taken the position that precise estimates of free-ridership cannot be measured in an era of comprehensive energy efficiency efforts at the local, state and national levels. In fact, some would argue that this issue of attribution is not important. In short, reasonable people disagree on how to separate net from gross savings, on how to account for indirect savings, and on how to parse savings to programs.

This paper will discuss how SMUD is re-visiting its measurement and evaluation protocols, in order to more accurately capture all of the benefits of their energy efficiency efforts. The results will be useful to evaluators and policymakers, at SMUD and elsewhere, who should re-think the traditional evaluation approaches and protocols, as well as to program planners who are challenged with ambitious new savings goals.

Introduction

The Sacramento Municipal Utility District (SMUD) has been a national leader in energy efficiency. In May, 2007, the SMUD Board set an ambitious 10-year goal to acquire energy efficiency resources equivalent to 15% of the forecasted system-wide energy use by 2017. This required a top-to-bottom re-design of SMUD's portfolio of energy efficiency programs, with expanded program offerings for all market sectors. In addition to adding new measures to rebate programs, the portfolio includes a new integrated marketing approach, expanded education and training, targeted customer behavioral changes, and codes and standards efforts.

With vastly more ambitious savings goals, and with a wider range of savings opportunities that will be targeted, it is important to revisit the question of how total savings and program savings are measured and evaluated.

Limitations of Traditional Approaches

Under SMUD's traditional program approach, savings have been relatively easy to measure. The emphasis was on discrete measures, with deemed or measurable savings, which could be counted and "banked". This approach is very defensible, and works well within the traditional utility program context. However, it has some serious limitations.

Traditional Approach Reviewed

To briefly review, the traditional measurement and evaluation approach in California typically, and at SMUD specifically, includes these steps:

1. Track direct utility efficiency interventions. This is typically done by counting customer participants in rebate programs; if a customer did not sign up for a rebate, his efficiency improvement was not counted.
2. Determine energy savings for the participant measures (only). Savings can be determined through direct measurement, calculation, deemed values, etc., and are measured in relation to a standard baseline. Savings are based either on standard assumptions or on the observed hours of operation in the post-installation situation.
3. Verify that measures have been installed and are operating as expected. Missing measures translate into missing savings.
4. Add up these savings to arrive at the gross savings estimate.
5. Estimate the amount of free-ridership. Free-riders are those customers who would have implemented the efficiency measures even without the direct utility incentive/rebate, and the savings from their measures are considered free-ridership.
6. (Optional) Estimate the amount of spillover, which is the saving from non-participant customers who implemented the efficiency measures because of utility program influence, and from participants who were influenced by their participation in the energy efficiency program to install more efficiency measures but who did not apply for or receive a utility incentive/rebate to do so. (Optional because not always considered as legitimate program savings.)
7. Subtract free-ridership savings from the gross program savings and, if applicable, add in the program savings from spillover, to arrive at net program savings. Calculate the net-to-gross ratio.
8. Count only the net savings toward meeting the savings goal.

Limitations

Each of these steps in the traditional approach can lead to missed savings that are not counted. Applying the same numbering as above:

1. Because only participant events are counted, indirect participation is not counted. For example, a customer may decide to drop out of the program for any number of reasons (paperwork hassle, program delays, etc.), but may still install the measure; the program may not be able to count those savings. A similar problem can occur with a customer who benefits from an upstream savings program, and so may not be properly identified or

counted. Even if counted, an upstream measure may over- or under-count actual savings if the final destination for the measure is not known.

2. The direct energy savings may not be correctly counted. If the adopted baseline does not correspond with reality (which can happen when energy code baselines rather than actual practice baselines are used), savings may be over- or under-estimated. Similarly, when occupant behavior or operational parameters are assumed rather than measured, savings can be incorrect. If occupant behavior or operational changes take place along with the measure installation, and savings are based on post-installation behavior/operation, then the savings due to the pre- and post-installation changes may be missed.
3. If the verification process misses installed measures, then savings are not counted.
4. Summations should be correct
5. Free-ridership calculations frequently under-estimate indirect program influences, especially when asking customers to self-report program influences. If such influences are assumed to be part of the naturally-occurring measure adoptions, they end up as free-ridership penalties. In other words, programs can be penalized for their own indirect savings. Also, for a mature program, a customer may have a history of participating and so no longer attribute their savings to the program.
6. Spillover can be difficult to measure, but if it is not counted, then savings are understated. Also, as noted in the preceding point, spillover or indirect program influence can turn into a free-ridership penalty depending on how attribution is measured.

Evaluation Problem Examples

The traditional approach has difficulty capturing less obvious savings. Some examples:

- A lighting retrofit project's energy savings are a direct function of the operating hours for the lighting system. The operating hours may simply be assumed based on the occupancy characteristics, or they may be measured on-site. Fewer operating hours mean fewer kilowatt-hours saved. The problem comes when the customer, after installing higher efficiency lighting, has also implemented operational changes to reduce the hours. This type of change is in the "low cost/no cost" category, and is the most cost effective way to save energy. But the traditional evaluation approach would calculate savings based on the observed (shortened) hours of operation, and the operational savings would not only not be counted, but the actual wattage-based savings during operating hours would be reduced. A thorough before/after measurement of total lighting energy uses would measure the full savings.
- An example on the spillover/net impacts side: by comparing per-household CFL sales in Massachusetts with per-household CFL sales in other (non-program) states, it has been estimated (conservatively) that the 2006 Massachusetts program was responsible for twice as many CFL sales as the program provided incentives for. The Massachusetts sponsors (through their evaluator) were using more conventional means to estimate savings per CFL, including tracking down older CFLs to estimate measure life, and using loggers to measure hours of use, etc. But counting the market effects and spillover gives them a big boost in measured savings—something they just claimed in 2007 for the first time.[1]
- Another example was given by a SMUD program representative. It concerned a chiller plant that was experiencing high chilled water loop pumping energy. The proposed solution was

[1] Evaluation studies in this case were conducted by HMG's project team member, Nexus Market Research. Anecdote reported by NMR Principal, Lynn Hoefgen (private communication 12/14/2007)

to install a variable speed drive (VSD) to more closely control the pumping energy. In reviewing the installation, however, the rep noticed that the valve on the chilled water bypass loop had been left open, so much of the pumping energy was simply sending the water in a loop within the equipment room. Closing the valve saved a substantial amount of energy at no cost, but it also substantially reduced the apparent savings from the VSD. The traditional evaluation protocol, based on ex post operation data, would miss the substantial savings from fixing the valve problem and would discount the savings from the installed equipment.

- At a broader level, the SMUD proposed portfolio redesign has recommended community-wide energy awareness, and urges a lot of no cost/low cost do-it-yourself savings within homes and businesses. These are legitimate savings targets, and the magnitude of savings could be large. The way savings are targeted and measured must be designed to capture the effects of these efforts; the traditional approach would miss them.
- Another example is codes and standards savings. In California, there is now a statewide protocol to estimate and credit energy savings from the 2005 building energy efficiency standards that the state adopted. Those savings are allocated to the IOUs, along with an attribution procedure to credit them for their efforts in adopting the standards. The savings allocations are based on utility sales within each service territory. Savings from these standards will show up in SMUD's territory, and will not be counted by any of the IOUs. The question is whether SMUD should count some or all of the savings within its territory. One answer is "no", because there is no attribution link between SMUD and the new standards. Another answer is "yes", because the savings are real. It is a SMUD portfolio policy decision whether or not to count these codes and standards savings.
- Similarly, SMUD customers can reasonably be expected to generate new energy savings due to spillover from the statewide Flex Your Power campaign, the Governor's directions to improve the efficiency of state buildings, the national ENERGY STAR marketing efforts, and neighboring PG&E's aggressive energy savings. These will be real savings that would not have occurred absent all these investments and efforts, but they are not going to be directly attributable to SMUD's program efforts. Should these savings be counted toward meeting SMUD's goals? If the answer is "no", then the savings will be ignored in calculating program savings, and may not even be counted toward SMUD's resource requirements. Again, one could answer "yes" because the savings are real. The decision is a policy choice for SMUD management and Board. The decision also bears on how those other programs count savings for their programs; do they ignore savings in SMUD's territory, or do they count them as well. Certainly there should not be double counting. This problem highlights the challenges we encounter when the regulatory policy is overly focused on attribution.

Discussion of Limitations

Of course, good evaluation practice and thorough program tracking can help to overcome these limitations. If the evaluation approach has limited time or resources, however, the simpler approach is to stick with the basics and count the obvious savings.

There may also be policy reasons to neglect counting some of the savings. For example, the California Public Utilities Commission (CPUC) has made a policy decision to ignore spillover in counting investor-owned utility (IOU) program savings. Evaluation protocols may also be deliberately conservative. The CPUC's evaluation protocols were established to ensure that efficiency savings are as real as new generation resources, and that the IOUs' accomplishments, and their shareholder rewards, were

conservatively measured and defensible within an adversarial public process among stakeholders. Other institutions are likely following these precedents as well.

SMUD, as a publicly-owned utility not subject to the CPUC's policy or political constraints, could adopt less conservative evaluation protocols and policies. Other program evaluators could do likewise. There are good reasons for doing this.

Effects of Evaluation Policies on Programs

Evaluation protocols have an important, and often unacknowledged, effect on how energy efficiency programs are designed and implemented. Good program managers will be running their programs to gain the maximum credit for their savings accomplishments, and will avoid spending resources on activities that do not result in credit. This is the well known "you get what you measure" effect. The converse is also true: "you don't get what you don't measure".

In many of the focus group discussions conducted for SMUD under the portfolio planning effort, program staff demonstrated a sophisticated understanding of how evaluation issues affect their program activities. While there may be good reasons that motivate program staff to pursue broader energy savings, the simple fact is that their goals, bonuses, career advancement and other incentives are driven by how savings are counted. When it becomes difficult to meet program goals, all efforts become directed at producing savings that will be counted (and at how those savings are credited). When program evaluation focuses narrowly on a limited definition of savings, and evaluation results are used to penalize program managers, then they naturally become less inclined to take risks and try new things, more defensive in how they operate their programs, and more restrictive in their rules for customer participation.

Thus, program policymakers and program evaluators must ensure that policies and evaluation methods align with the goals that are set.

Expanding the Evaluation Methods

Because SMUD has expanded its goals and is targeting much broader savings than in the past, the evaluation methods also must be revisited. SMUD is not under the same regulatory constraints as the IOUs, and so is thinking long and hard about whether to constrain itself by the same, relatively narrow evaluation protocols that govern the IOU programs..

One way to think about this problem is to ask the question, "Where do we want to draw the boundaries of program influence and savings?" The question relates to gross savings, to free-ridership to spillover and to market effects. We take it as a given that measured savings must be real, but as we've pointed out, there are different interpretations of "real". The answer will strongly influence how SMUD staff operate their programs, how SMUD tells its constituents about the results of the new portfolio, and how SMUD reports its savings to the CEC for statewide counting purposes.

Single Program Evaluation

Conceptually, one can think of the savings from a single program as a series of concentric circles, as shown in Figure 1. The direct participant savings are shown in the shaded circle, with the savings from free riders subtracted. There are additional potential savings that can result from spillover, both from participants and from non-participants.

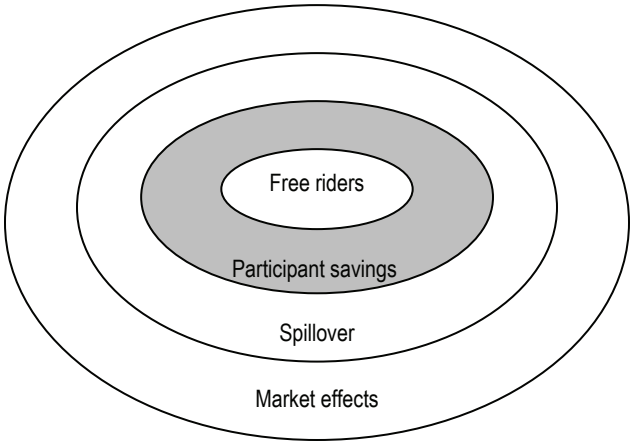

Figure 1. Single Program Evaluation Components

Additional program savings may be gained from wider market effects caused, for example, by merchants taking less efficient equipment off their shelves and selling only the efficient equipment to customers, with or without program incentives. These market effects may ultimately indicate market transformation and signal the time to remove program incentives. When measures are not yet widely adopted, however, market effects savings can represent real and cost-effective savings that the programs may encourage. Of course, this assumes that the program managers can expect the additional savings to be counted.

This single program approach has been widely applied in past program evaluations. It implicitly assumes that there is a distinct boundary that can be drawn around the program and its market segment.

Evaluators Dilemma - Multiple Overlapping Programs

An increasingly prevalent problem for evaluators is arising as energy efficiency programs become more aggressive. SMUD inevitably will face this problem as it pursues its ambitious savings goals. This is the problem of overlapping programs, as shown diagrammatically in Figure 2. The program influences spread out and overlap like the ripples from raindrops on a pond. This can occur for many reasons.

- Programs are penetrating more deeply into markets, and they are achieving substantial market shares.
- Programs do not always have clear boundaries within the broader market. Customer groups may overlap, and aggressive program managers may be approaching customers from several directions.
- Programs may be addressing both upstream and downstream market actors.
- Prescriptive and comprehensive program offerings may be covering the same measures, and so the same suppliers, but with different program treatments
- Multiple program administrators (e.g. neighboring utilities, statewide programs, national programs) may be focusing on the same measures and the same customers.

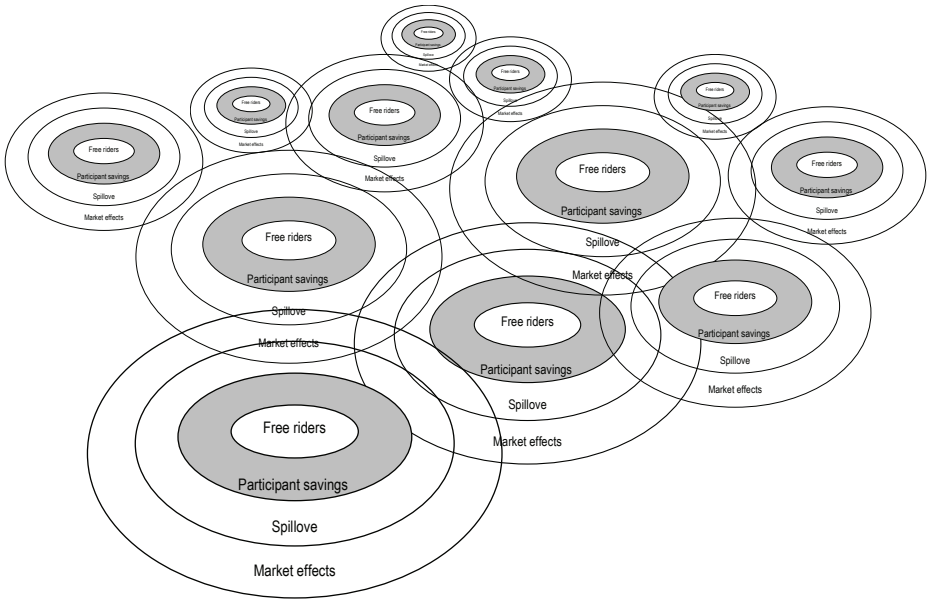

Figure 2. Multiple Program Evaluation Overlaps

When there are multiple programs operating concurrently, it becomes increasingly difficult to draw neat boundaries around each program, and to evaluate each independently. Certainly the programs within a given market segment will interact at the market level, but other interactions are also likely. The diagram shows the program participant savings as free of overlap, but the overlaps can even occur down to that level.

SMUD will experience a further dimension to this problem, because customers within its electric service territory are served natural gas by PG&E, which offers its own gas efficiency programs to those customers. In addition, SMUD's relatively small service territory is surrounded by other electric utilities, and it operates within the reach of California's statewide energy efficiency efforts, such as the Flex Your Power marketing program. There are also national programs, such as ENERGY STAR that target consumers. Program influences on customers are legion.

When there is this much overlap between programs and between customer offerings, it becomes increasingly difficult to measure free ridership, spillover and net participant savings. It can even be a challenge to measure gross program savings. Because of the multiple program overlap problem, evaluators may be forced to abandon efforts to evaluate each program offering separately, instead working to evaluate combined savings efforts at the level of major market segments. This would be consistent with recommendations to remove the "stovepiping" of individual programs, and to instead deliver integrated program offerings that can address all efficiency opportunities for customers within market segments. Stovepiping, the tendency for programs and their offerings to act as independent domains with separate rules, marketing approaches, incentives and managers, has been a problem at SMUD and with many other program portfolios as well. It leads to competition or lack of cooperation between program managers, confusion for customers, and lost opportunities when measures or customers fall between the stovepipes. Stovepiping can be a rational response by program planners to management or policy directives, but it is hardly the best way to run programs. Portfolio policies, the evaluation approach and the program design should be aligned; integrated evaluation is needed for integrated programs, and vice versa.

Evaluation Debate Beyond SMUD

These evaluation and program issues are hotly debated between the CPUC and the IOUs, in other jurisdictions, and within the evaluation community, driven primarily by the need to demonstrate real and reliable savings, but also largely influenced (for IOUs) by the substantial shareholder incentive dollars that

are at stake. It has been pointed out that the gross savings are what really matter to system planners and procurement needs; that net savings are only needed for shareholder earnings determinations and for improving program implementation. Many have argued that the CPUC's determination to penalize for free-riders, while not recognizing free-drivers or spillover, is asymmetric and unfairly punitive for program administrators. Other regulators have recognized spillover and market effects (e.g. New York). People have also argued that it is nearly impossible to accurately measure free-ridership in an era of comprehensive energy efficiency efforts at the local, state and national levels. California probably spends more on evaluation efforts focused on parsing out these issues than any other jurisdiction in the US. As a comparison, evaluators and policymakers in the Pacific Northwest focus their efforts on accurately measuring gross savings, and frequently arrive at net savings allocations by negotiation, rather than by extensive measurement efforts. Reasonable people disagree on how much emphasis to place on separating out net from gross savings, on measuring spillover and market effects, and on the need for (or feasibility of) precisely measuring free-ridership.

Evaluation Policy Recommendations for SMUD

In setting its ambitious savings targets, the SMUD Board did not explicitly rule on these evaluation issues. They left some latitude, therefore, to look at them more closely, and to recommend a less-traditional approach.

The recommendations to the SMUD Board and management, from SMUD staff and their consultants are:

1. Set the overall goals for *gross* savings from all credible methods of saving energy, both physical and operational. Gross savings can be defined as including any energy savings that otherwise would not have happened, absent all SMUD efforts. A technical way to define gross savings is to compare total electrical energy use to a reasonable baseline level of consumption (or to the "before" energy use when it can be measured, as in a large retrofit).
2. Encourage rigorous but innovative methods for determining savings for both traditional and non-traditional energy savings across all end-uses and customer classes. Encourage and recognize operational energy savings, and encourage changes in behavior that provide persistent savings. In other words, do not limit savings to hardware changes, such as reductions in installed lighting wattage or improvements in air conditioner EERs, as has been the past practice.
3. Measure free ridership and spillover, but at the market segment level, rather than program-by-program, so as to capture all utility influences and avoid program overlap problems. Use the resulting net-to-gross ratio(s) primarily to inform forward-looking refinements and realignments to program activities (as opposed to assigning failure or blame, penalties or rewards).
4. Rely on process evaluations and market studies to gather *timely* information for making effective mid-course corrections in program strategies and offerings, and to assess the degree of market transformation that has occurred. Err on the side of doing these studies earlier rather than later.
5. Avoid warping SMUD's measurement and evaluation efforts by NOT directly copying the CPUC/IOU shareholder earnings-based approach to M&E.
6. Establish mechanisms for evaluating program offerings that integrate multiple measures of success, rather than relying on simple, numerical savings goals. Such mechanisms should recognize program lifecycles (from pilot to mature to phasing out). They should also acknowledge efforts to capture lost opportunities, to build future capacity, and to produce permanent changes to customer attitudes and behaviors.

7. Develop staff performance goals and incentives that are comprehensive in nature. Avoid stovepiping of programs by avoiding simple numerical targets at the program level, which can become incentives for cream skimming and/or competition between programs for customers and which may result in customer confusion.

Conclusion

The issues that SMUD is facing, of aligning their program goals with their evaluation approaches, are universal. Policymakers, program designers and implementers, and evaluators should all be aware of these issues, and should consider how they apply to their own program portfolios. Aggressive savings goals require the pursuit of all available savings opportunities, which in turn requires aggressive evaluation approaches to capture all of the resulting savings. If evaluation policies limit the savings that are counted, they will, in turn, limit the amount of savings that will be realized. This paper has reported on the thinking that is informing SMUD's program plans for the next ten years, and it is hoped that this will contribute positively toward advancing the fields of energy efficiency and evaluation.

Acknowledgements

The authors wish to thank the management and staff of the Sacramento Municipal Utility District for its support of this work. In particular, we acknowledge the contributions of Bruce Ceniceros, Janet Erickson and the rest of the program and evaluation planning staff who guided the project and supported the exploration of new and interesting program and evaluation approaches. We also acknowledge the members of the consultant team: the Heschong Mahone Group Inc., lead by Douglas Mahone, Cynthia Austin, Lisa Heschong, and others; and the subcontractors: Quantec LLC; Nexus Market Research; Itron; and Steve Schiller.

References

Heschong Mahone Group Inc. "Energy Efficiency Portfolio Redesign - Developing a New Plan for SMUD's Efficiency Programs." Report prepared for the Sacramento Municipal Utility District. July 28, 2008.