Pursuit of Aggressive Energy Savings Targets: Aligning Program Design and Evaluation

Douglas Mahone, Heschong Mahone Group, Inc.
Nick Hall, TecMarket Works

ABSTRACT

This paper discusses some of the current worst and best practices for coordinating program design and evaluation efforts to address the challenges of aggressive energy efficiency. It makes the case that evaluation considerations, and the expertise of evaluators, should be aligned with portfolio long-term goals, and that program managers should work with evaluators to capture the full savings potential of the programs, and of the portfolio. This paper also suggests that policymakers set policies for both program design and for applying evaluation protocols to establish a timely and rational evaluation framework that also helps to encourage maximum savings capture. The paper draws on the experiences of the authors in both program design and evaluation in different regions of the US to illustrate these points. The paper deliberately does not 'name names', in the interest of identifying and discussing broadly applicable issues and recommendations.

Background

When a program manager implements a standard type of energy efficiency program, with a set of energy efficiency measures typical for that type of program, the program design and the subsequent evaluation effort to measure its savings are also well understood and fairly simple to implement. The traditional measurement and evaluation approaches suffice. In most states, these studies repeat the approaches and analysis frameworks that have proven to be reliable over the last 35 years of energy efficiency portfolio and program evaluation.

It's not so simple, however, when portfolio managers are charged with pursuing aggressive short-term and long-term energy savings targets across multiple markets and market segments, with many more efficiency measures. They may find that some programs compete with the objectives, or minimize the accomplishments of other programs within the portfolio and within the larger market.[1] The challenge is not only to capture as much savings as possible from the usual energy efficiency measures, but also to pursue non-traditional energy savings that go deeper and last longer. These may include the pursuit of upstream savings from manufacturers or distributors, multi-year retrofit strategies, behavioral changes to reduce energy use, market transformation savings resulting from training or changes in design practices, non-standard long life measures, savings from enhanced codes and standards, and other innovative strategies. All of this creates new challenges in both program design and program evaluation.

We focus the current discussion on entities that are charged with generating energy savings (and demand reductions) across a broad spectrum of market sectors and technologies. Most typically these are utilities, but they also include other governmental, non-profit and private

[1] Of course, good program and portfolio design can minimize many of these problems, but they are unfortunately quite common.

sector entities. We will refer to these people as portfolio managers, in reference to the portfolio of energy efficiency programs they design and implement.

The policies that portfolio managers implement are typically set by governments, their policymakers, utility boards of directors, or other entities charged with safeguarding the public good (herein referred to as policymakers) by assuring the uninterrupted supply of least cost energy resources (including energy efficiency), counting not only the cost of the generation and distribution, but also the costs associated with using that energy.

Together and collectively, portfolio managers and policymakers are engaged in a grand and complex enterprise with many interrelated and sometimes contradictory parts, involving markets, manufacturers, suppliers, contractors, building owners, occupants and other market actors. They are also doing most of this work using OPM (other peoples' money), in the form of ratepayer or taxpayer dollars. This means that there is, naturally, a great deal of scrutiny to ensure that the dollars are well spent for the intended purpose of providing least total cost energy supplies, and that the energy efficiency "supplies" are real. Real energy efficiency induced savings typically save dollars, reduce energy consumption and greenhouse gases, avoid the need to build power plants or other forms of energy supply, and mitigate the risks and costs associated with all of that.

To do this, policymakers set energy efficiency goals for the portfolio managers to achieve, using the available funds to overcome market barriers, and to help building owners and other customers implement energy efficiency improvements in their buildings and energy using equipment. A key consequence of all this is that policymakers need assurances that energy savings go beyond whatever energy efficiency improvements would have naturally occurred in the marketplace (else, why pay for them?). They also want assurances that the value of the energy savings is commensurate with the resources spent to acquire them. Finally, they want to know whether the portfolio managers met their savings goals, as there are often incentives and penalties tied to those goals.

This is where evaluators, and evaluation protocols, enter the picture. It is their job to measure and evaluate the outcomes of the portfolio managers' programmatic efforts, providing expert and independent review of the outcomes. The evaluators are asked to verify that energy efficiency measures have been installed as promised, and to calculate the gross energy savings from those measures. These are then compared to the portfolio managers' estimates of gross savings to achieve a realization rate. The evaluators are also asked to estimate what share of those savings would have occurred without the programs' efforts; program participants who would have acted anyway are deemed to be free-riders, and their savings are subtracted from the gross savings to arrive at the net program savings. The answer to this calculation is different when the evaluators are asked to estimate savings net of the *portfolio's* efforts. This is especially true when the portfolio includes activities that erode net *program* savings, but increase net *portfolio* savings (e.g. by increasing apparent program free-ridership while improving market penetration). Unfortunately, policymakers seldom ask their evaluation contractors to design their studies to assess total net *portfolio* savings. These are different types of studies that focus on different populations and different cause-and-effect relationships, which are typically not recognized within an individual program's program theory. Instead policymakers ask their evaluation contactors to add up the program-specific savings estimates as if the summation of the program-level savings represent the net total portfolio savings.

Often, the evaluators are also asked to go farther and estimate the savings that occurred as a result of spillover from the program or from broader market effects that can be traced back to program influences. Whether or not this is the case, policymakers adopt evaluation protocols that spell out the scope of the evaluators' tasks, and reflect the policy choices made by the policymakers in deciding what to credit and what to ignore when counting up savings.

With this brief background on portfolios and evaluation, let us now look at the current state of many energy efficiency portfolios in the US. According to a 2009 ACEEE study:

"A number of states have set savings goals for utility-sector energy efficiency in the range of 1.5% - 2.0% of total sales each year." These are aggressive goals, considering that only one state (Vermont) had yet to achieve savings higher than 1.5% per year[2]. [ACEEE]

The study points out that even the most prominent energy efficiency portfolios have only achieved savings in the 0.5% - 0.8% range of total sales per year, with only three higher exceptions. Whether or not the new, more aggressive goals are reasonable, it is clear that a dramatic ramp-up in energy efficiency programs will be required even to come close to meeting these savings goals. In order to double or triple the level of energy savings, will not only require more dollars, but will require an expansion in the scope and reach of efficiency portfolios.

Traditional energy efficiency programs have focused on simple, transaction-based efficiency measures, such as providing rebates for specific retrofit measures (air conditioners, furnaces, lighting, chillers, boilers, pumps, motors, compressors, insulation, weatherization, etc.). These are often referred to as the 'low-hanging fruit', because they are can be acquired at substantially less than the avoided cost of a new traditional supply, are relatively easy to implement, the equipment and technologies are readily available in the marketplace, and the savings are easy to measure. The efficiency program transaction is straightforward: participants are invited to participate, they are offered a rebate or other assistance in installing the measure, and the program books the savings. The measures are also straightforward to buy and install, participation rates can be relatively high because the measures are not difficult or unreliable, and program costs can be low. However, unless the program is structured to weed them out, there can also be a high level of free-ridership because of all this.

To meet the aggressive new savings goals will require a broader range of strategies. These may include:

- Upstream initiatives offering subsidies to manufacturers or distributors
- Multi-year or comprehensive retrofit strategies with customer organizations
- Behavioral change initiatives to reduce energy use by building operators and occupants
- Market transformation initiatives resulting from training or changes in design practices
- Savings from enhanced codes and standards
- National or statewide outreach and awareness programs
- Changes to existing public policies to require energy efficiency and climate change measures to be pursued ahead of traditional carbon-based energy supplies

Higher levels of savings can be captured from the interactive effects and synergies between and among these program strategies and policies. All of this creates new challenges in both program design and program evaluation. Under a typical program or portfolio evaluation framework, especially when more innovative, aggressive, deeply integrated and comprehensive

[2] Note that "aggressive" can be a relative term. Goals of 1%-2% may not even reduce load growth. Some do not apply the term "aggressive" until savings targets exceed 5%

strategies are required, the traditional methods of booking and measuring savings begin to fall short. They may credit only parts of the portfolio (incentive-based resource programs). The kind of assessment these broader program efforts deserve, and the identification of free-riders and spillover becomes more difficult. These challenges to program designers and to evaluators, and the policy changes that will be needed to guide their efforts toward higher goals, are the subject of this paper.

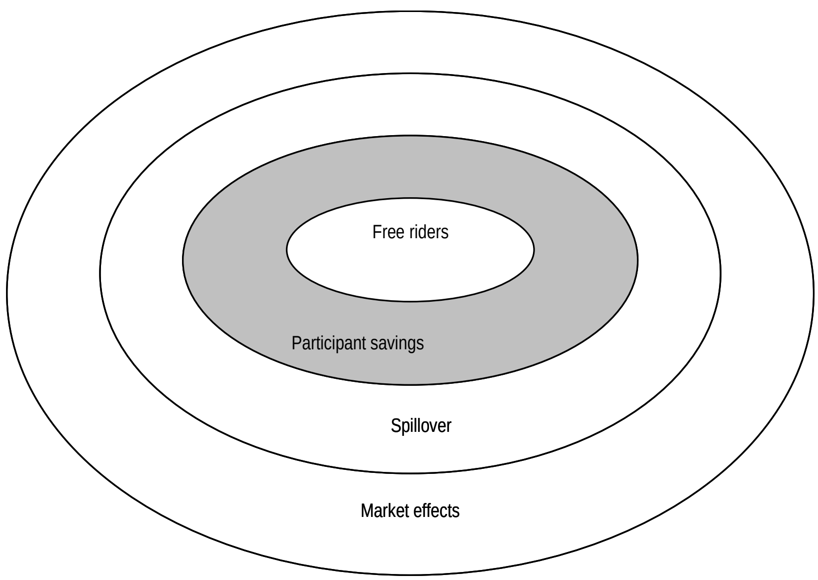
In preparing this paper, the authors have deliberately avoided 'naming names' or singling our specific programs, portfolios or policies. This is to avoid arguments about specific portfolio and policy settings, and instead to focus on the problems presented by traditional program and evaluation practices that can be avoided in the pursuit of aggressive energy efficiency.

Evaluators' Dilemma

We start by explaining the evaluators' dilemma, because in many ways the results of energy efficiency portfolios are determined not by what is actually achieved but by how the savings are measured, i.e., you get what you measure.

The basic evaluation approach for traditional programs can be simply described by the following diagram.

Figure 1. Single Program Evaluation Components



The participant savings, which are the verified savings from the efficiency measures installed through the program transaction with the customer, are shown in the shaded circle. Any of those savings attributable to free riders are subtracted out (inner circle), leaving the shaded donut of net participant savings. If there are spillover savings, either from additional measures installed by participants without program incentives, or from measures installed by non-participants who were indirectly influenced by participants' actions, these may be added to the total. Finally, as the programs begin to affect the market, e.g. by inducing retailers to sell only efficient equipment in response to market demand, there may be additional savings, called market effects. These additional savings would expand the shaded area of countable program savings.

Some policymakers have limited the savings that can be counted as net program savings, in an effort to maintain the rigor and reliability of the savings estimates. Some of the policy rules that have had this effect include:

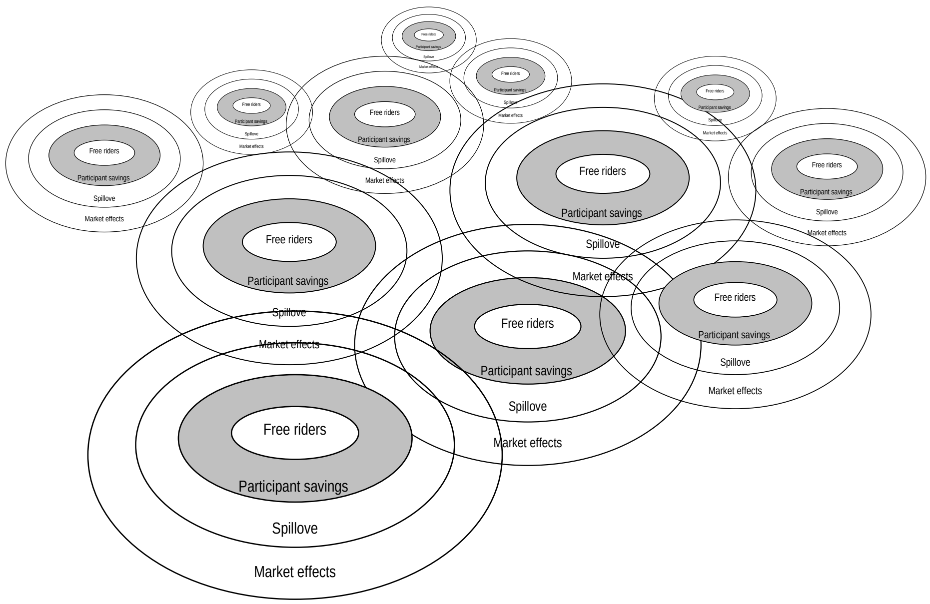
- Only count savings when it is possible to attribute a direct link between program influence and participant action
- Do not count spillover, or only count clearly identified participant spillover
- Do not count market effects
- Only count measures that are recruited and installed within a program cycle
- Do not count participants who were influenced by other efficiency programs
- Do not count participants who participated in, and so were influenced by, earlier program cycles

As can be clearly seen by the above points, the rules of the game can be established to limit recognition of program efficiency accomplishments. While policies may proclaim that energy efficiency is the resource of first choice, the policies within the evaluation framework can work to minimize the accomplishment of this objective. This would be like counting the full cost of a new power plant on the cost side of the equation, but only allowing the generation occurring on Saturdays and Thursdays to be counted as a benefit.

This traditional approach to evaluation protocols and practices has been satisfactory for traditional efficiency program efforts, especially when programs were targeted to discrete measures and customer groups. This approach also encourages portfolio managers to implement programs with discrete boundaries and clear lines of attribution; else their program savings will not be counted. They are not encouraged to think or plan in terms of broader energy policies, or to consider how best to reach state or national carbon reduction objectives.

In areas with aggressive efficiency program goals, the programs almost unavoidably overlap. Rather than the simple evaluation of discrete programs, as shown in Figure 1, evaluators are faced with portfolio situations such as shown in Figure 2. This is the result of more program offerings to virtually all customer segments. It is also the result of policies directing programs to go beyond the low hanging fruit and to find deeper savings. There is also a time dimension to these overlaps; as programs extend over multiple year time periods, and as they seek savings that may take several years from initiation to completion, current programs will affect future programs and will be affected by past programs. Of course, Figure 2 is not even a worst case illustration, because none of the participant savings are shown to overlap (for graphic clarity). In reality, there could be significant overlaps, with one program's participants being another program's free-riders.

Figure 2. Multiple Program Evaluation Overlaps



Faced with these overlapping programs, evaluators have an increasingly complex task of parsing out program accomplishments and attribution, and of measuring net savings for each individual program. For example, if a customer participates in a retrofit rebate program at the same time there are national and statewide information programs, tax credits, and technical assistance offerings, and if the customer had participated in a related program three years prior, is the savings from the retrofit measure really due to the current rebate program? Should that customer be deemed a free-rider? It becomes increasingly complex to answer these questions. Even if the evaluator had access to all the program records from all the related program activities, and could document all of the influences on the current retrofit, it would not be clear how to interpret the interrelated chain of influences and thereby arrive at a decision on attribution and credit. It is clear, however, that if all the policy rules listed above for a standalone program were applied in this context, the net savings attributable to the retrofit rebate would be low. By extension, the savings not counted for a portfolio across all programs, would be large. The very nature of activities such as marketing, outreach, education, information, training, and the creation of new business dynamics should result in long-term behavior change and market effects. These program activities can actually end up penalizing the portfolio and program savings. The constituent parts of the portfolio end up cannibalizing each others' savings because of the policies embedded within the traditional evaluation framework and the way policymakers credit savings. Essentially this threatens failure of the portfolio to achieve longer-term state and national energy and carbon objectives because of the policies embedded in the traditional evaluation frameworks and the approach for crediting impacts.

The question that evaluators receive from policymakers is: "where do we draw the program boundaries, and should we even try to separate out savings and attribution at the level of individual programs?" Some suggest that portfolio managers should be creating *combined* programs, say around all the program offerings to an individual market segment. Let us explore the implications of all this for the portfolio managers and their program evaluators.

Program Managers' Dilemma

Portfolio and program managers have been given a challenging task in the design and implementation of energy efficiency programs. They must balance a number of competing objectives:

- **Meet goal -** Focus on capturing energy savings that meet or exceed established goals within the immediate program timeframe
- **Provide effective incentives** - Spend enough money to attract participants and motivate efficiency actions
- **Be cost efficient**- Spend only as much money as necessary to achieve the goals
- **Work with the market** - Partner with other market actors, recognizing their needs and desires, to effectively change efficiency practices and products in the market
- **Avoid free-riders** - Figure out how to identify them, avoid wasting program resources on them
- **Comply with policy directives** - Even if they conflict or are not internally consistent
- **Satisfy management** - Whether they be shareholders, constituents or policymakers
- **Meet both short- and long-term goals** - Meet goal, but encourage long-term spillover and market transformation
- **Keep staff satisfied and growing** - This is their career, and their success depends on the program success, and also on how well their day-to-day work life is managed
- **Etc.** - the list goes on…

Of course, an efficient manager will seek to simplify all of this as much as possible, to keep it manageable. This can lead to an emphasis on the specific, short term objectives at the expense of the broader, long-term objectives. If measurement and evaluation protocols also focus on the specific, short-term objectives, and do not count the results of efforts to address the long-term objectives, then it is only natural that program activities are narrowly defined and implemented. This is one of the causes of the 'stovepiping' of programs (when programs only focus on a narrow market effect or segment and do not integrate with other program offerings). In the most extreme cases, programs directly compete for customers at the expense of other programs. From a portfolio perspective, this can be damaging, and from a customer perspective it's confusing. Broader goals of deep energy savings, market transformation, multi-year customer participation and integrated program offerings are abandoned because the savings may not be counted.

It is a well-known principle of management that 'you get what you measure.' The corollary is that 'you do not get what you do not measure.' Program designers and implementers are quite sophisticated in their understanding of how their efforts will be evaluated. When evaluation protocols are not clear or when there are contradictory policies reflected in those protocols, then program managers are faced with 'evaluation risk,' the possibility that their efforts and results will not be credited to their programs (examples discussed below). Any good manager will seek to manage and reduce that risk. In our diagram of overlapping programs, one way to manage that risk is to focus program efforts as narrowly as possible on results that have the least probability of being judged free-riders, or of overlapping with other program efforts (including past efforts).

The following section describes examples from recent program experience that illustrate these problems. For simplicity, we will refer to narrowly-focused, strictly applied evaluation practices as “traditional evaluation protocols”. In doing this, we are describing broadly applied, though by no means universal, evaluation practices

Examples of Program Design and Evaluation Pitfalls

The Case of the Ubiquitous CFL

Ten years ago, compact fluorescent lamps (CFLs) were expensive and rare. Utilities, the federal ENERGY STAR program, retailers, and consumer advocates have been pushing CFLs for years, seeking greater market acceptance and lower costs. Some program managers, faced with large savings goals and short time periods to perform, have focused a great deal of program resources on encouraging this to happen. They have offered up-stream incentives to manufactures, distributors and major retailers to make mass quantities of CFLs available in the market at reduced cost. They have offered direct incentives to consumers to purchase the bulbs. They have bombarded the public with advertising and public service messages. In all of this, they have largely been successful. Evaluators have found that CFLs have gained widespread acceptance, and in many parts of the country have found that they are readily available on the shelves alongside traditional incandescent light bulbs. The problem, for program managers, is that evaluators are also finding evidence that the market for CFLs has been transformed, at least partially. Customers who were once reluctant to pay extra for CFLs have now accepted them and are buying them for at least a part of their lighting needs.

The question, then, is whether current program efforts to encourage the use of CFLs are now all directed at free-riders, and are therefore wasteful and ineffective? Or is this a sign of long-term program success and market transformation that should be rewarded? Traditional evaluation frameworks can convert a market success metric into a program failure metric by not recognizing the success over time. This is not to suggest that the CFL market is fully transformed, or that substantial continued efforts are not needed to continue that market success. In fact, as noted by the U.S. EPA, the market is still largely untransformed, sales of standard light bulbs still outpace CFL sales, and the vast majority of light sockets are still fitted with standard bulbs. But this example does show how different evaluation policies and protocols will lead to different definitions of success and achievement in energy efficiency portfolios.

The Case of the Enlightened Institutional Energy Policy

Some forward-looking efficiency program managers have devoted program dollars to working with major customers (corporations, community colleges districts, hospital groups, etc.) to help them adopt institution-wide policies to seek energy efficiencies wherever possible, and to work with their local energy efficiency organization to maximize energy savings. Some of those policies have been in place for more several years. Then along comes the energy efficiency program implementer offering rebates and technical assistance to help the institution adopt the latest energy efficiency measures. The measures are installed, the rebates paid, the savings verified. When the evaluators, applying traditional evaluation framework requirements, determine program free-ridership, they ask the energy managers at the institution whether they would have implemented the efficiency measures absent the incentive program. If they are told, "*Yes, certainly, because it is our established policy to maximize energy efficiency.,"* the evaluator must disallow the saving achieved by the program in accordance with the net-to-gross policies established in the traditional evaluation framework.

If the focus of the evaluators is only on the most recent transactions, they will miss the historic relationship between the efficiency organization and the institution(s) that led to those savings. The transactions are judged to be free-riders in the current efficiency program, and the savings are not counted. The program managers are judged to have wasted program resources, the institution is judged to be gaming the program, the portfolio goals may not be met, and the portfolio managers have shot themselves in the foot because they chose to take the long view on deeper savings in prior program years. The evaluators have done their job, as directed by the traditional protocols. However, the policymakers have failed, because while they may say they want long-term market transformation, they have set policies that militate against it.

The Case of the Missing Operating Hours

One of the goals of most energy efficiency programs is to first encourage 'no-cost, low-cost' measures to save energy, before investing dollars in new equipment or systems. If a program persuades a customer to make substantial changes to a lighting system by installing new fixtures, lamps, ballasts, and controls, there will be energy savings that can be measured in kilowatts. To convert these into energy savings, it is necessary to multiply the kilowatt savings by the number of hours of operation. This is typically measured by evaluators upon completion of the project, based on observed operating practices. The problem arises if the operators or occupants implemented the no-cost measure of reducing the operating hours, while the new lighting system was being installed. The evaluators, following the traditional evaluation framework, will only measure the observed (reduced) hours of operation, and will miss the full savings achieved. Furthermore, this will diminish the energy savings attributed the retrofit, so there is a double ding to the program savings. A clever program manager would counsel the customer to make no changes in operating hours until after the evaluators had done their measurements. A better solution would be to measure the operating hours before and after the new lighting system, so as to capture both the kilowatt reduction and the operating improvement. Again, it is a matter of evaluation policy and protocol whether this is done. The problem applies to all kinds of no-cost, low-cost actions taken by customers, including measures that deal with fans, motors, cooling towers, air-conditioners, and other measures where operating hours are a key variable in the savings.

The Case of the Missing Spillover

In a residential CFL (compact fluorescent lamp) program, the traditional evaluation approach is to go into homes, looking for CFL installs, measuring operating hours, looking for signs of persistence (older CFLs). A more advanced evaluation practice is to look at the entire market, by comparing CFL sales per household in the program area to sales per household in a nearby non-program area. In some of these evaluations, sales were shown to be twice as high in the program area, compared to the non-program area. In an even more advanced evaluation framework, the evaluator would look for effect of the various programs on national and regional markets, and would identify how programs have induced savings in the program area, in the utility territory, in the state, in market regions and at the national level. However, traditional evaluation frameworks allow only the counting of a small part of the total effects. This is a strong indication of spillover savings that would not be captured by the traditional evaluation. While spillover can be difficult to measure, this example illustrates the potential of substantially undercounting savings if spillover is ignored.

The Case of the Mistimed Savings

New construction projects can take years to be completed (especially in a down economy). It is not unusual for it to take three years from the time a new building design is started until the building is constructed and occupied, and the savings become real. This is also true of many other large energy efficiency projects, such as industrial or central plant upgrades. This fact can come into conflict with program funding cycles and the accountability for meeting savings goals within those cycles. In some jurisdictions, savings are 'booked' when a participant project signs onto a new construction program. The project may not be completed, and the savings may not actually show up, for several years, and program savings may need to be retroactively adjusted downward if the project never completes or if it abandons its efficiency measures. In other jurisdictions, savings are not 'booked' until the new building is completed and comes 'online'. In that case, some projects signed up in a given program cycle may only produce savings that are counted in a subsequent program cycle. In both cases, however, the energy savings are recognized, and the long lead time for the savings are accommodated.

If, however, a jurisdiction only counts savings that are both initiated and realized within a discrete program cycle, then many projects will fall through the cracks and never be counted. For example, if the program cycle runs for three years, any three-year new construction project would have to be initiated during the first months of the program cycle, to ensure that it was completed and on-line before the program cycle ends. A program manager who devotes program resources to working with a three-year project in the second or third year of the program cycle would be wasting time and money, because the energy savings would not come online until after the cycle had ended, and so the savings would not count. Likewise, any customer unlucky to be starting a major project that couldn't be completed before the end of the cycle would be ineligible for program assistance. Only those customers lucky enough to fit neatly into the program cycle could benefit from the program. A policy of only counting savings that are both initiated and realized within a given program cycle, therefore, ignores large savings opportunities.

The Case of the Misapplied Best Practices

There have been numerous efforts to identify best practices in program design and implementation. For example, in 2004, a joint effort between the California Investor-Owned utilities, the California Energy Commission, and the California Public Utility Commissions produced a set of best practices [CPUC] to identify and communicate excellent program design and implementation practices found nationwide, in order to enhance the design of energy efficiency programs in California. In particular, program implementers, supported through public goods funds in California, were encouraged to use this Study's products, along with other resources and their own knowledge and experience, to develop and refine energy efficiency programs. Some of the recommendations included the following:

- Use the ENERGY STAR logo to instill consumer confidence
- Provide trade allies with training & resources to enhance efficiency marketing
- Sell the customer benefits first, then sell energy efficiency

The first two recommendations follow the strategy to leverage various channels to reach the customer through federal programs and market actors. However, if the traditional evaluation policy only counts savings from customers who were solely influenced by direct program marketing materials, then leveraging marketing channels would eventually "count against" a program - despite following best practices recommendations.

Similarly, in the third bullet, for customer efficiency measure adoptions, traditional evaluations might treat those customers influenced by non-energy benefits as free-riders.

The Case of the Self-Defeating Cost-Effectiveness Test

Policymakers naturally want assurances that efficiency dollars are spent on measures that are cost effective, and so do not waste money on efficiency if there are better ways to use the resources. Typically, a key measure of cost effectiveness is the Total Resource Cost (TRC) test. This test compares the benefits of energy efficiency against all of the costs. The benefits are calculated as a function of the avoided cost of energy, a discount rate to account for the time value of money, an effective useful life for the analysis, and possibly the value of avoided carbon or other environmental benefits. If these parameters are set too conservatively, then the only measures that will be cost effective will be the "low-hanging fruit", which are obviously cost effective. But longer term, deeper energy savings will necessarily be less cost effective, and may even be ruled out by the TRC, if the parameters are not set higher. For example, many jurisdictions set the avoided cost of energy for efficiency measures at a level equivalent to power costs from a new, combined-cycle gas turbine generator. Some of these jurisdictions, at the same time, set ambitious goals for renewable energy investments, which produce energy for a much higher cost than a gas turbine generator. It would make sense, in such a situation, to set the avoided cost for efficiency measures to the higher cost for the renewable portfolio. Similar comments can be made about the consequences of setting the discount rate too high, setting artificially short useful life assumptions, or undervaluing the environmental benefits from efficiency. On top of that, some policymakers require efficiency programs to show TRC benefit/cost values that are substantially higher than 1.0. This means that measures must be even more cost advantageous than the TRC parameters have specified. While there is nothing wrong with wanting highly cost-effective savings, setting TRC parameters low and targets high has the effect of ruling out a lot of efficiency measures that would actually be cost effective. Under these circumstances, deep energy efficiency and ambitious savings goals cannot be achieved.

Policymakers' Conundrum

Policymakers and regulators, of course, sit at the nexus of these problems. They set the goals and get the energy efficiency programs going. They set the policies for how programs will be implemented. They adopt the policies and evaluation protocols for how program savings will be counted. They establish the rewards and penalties for program implementers. And they are accountable to their ratepayers or constituents for how the energy efficiency dollars are spent.

When policymakers set especially aggressive savings goals, all of this becomes more acute. The portfolio managers are challenged to go after every potential savings opportunity, to address every customer and industry, and to use all available strategies. The efforts will necessarily overlap. Indeed, overlap and inter-program synergies can be part of an overarching strategy, whether the goal is controlling load growth, reducing greenhouse gas emissions, or transforming markets.

The problem, as we have been illustrating, is that the need for strict and conservative savings accounting (traditional evaluation) comes into conflict with the need to aggressively pursue every potential savings opportunity. If the acceptable savings are defined too narrowly, then the programs will perforce become too narrow to meet the aggressive goals. If the savings are counted too broadly, then program dollars may not be prudently spent, or accomplishments may be overstated.

This effect also impacts program cost-effectiveness. If only the most narrowly-defined savings are counted toward the net savings, then more program resources will be counted as having gone to free-rider participants and overall cost-effectiveness will suffer. Yet policymakers do not want to count every kilowatt-hour claimed by program implementers, because not all of the savings will be 'real'. This also adversely affects cost-effectiveness.

Policymakers report that they are caught in a policy catch-22. Policymakers are often governed by enabling law that requires them to focus their policy rules on least cost reliable energy supplies; a noble objective. However, there is a policy clash between least cost energy supplies and their ambitious energy efficiency goals (and with national or state efficiency and carbon reduction objectives). Policymakers will say that they are placing themselves out on a limb and sawing off the branch if they increase the cost of energy in order to reach their efficiency or carbon reduction objectives.

Renewable Energy Standards that require the construction of renewable energy facilities have provided policymakers with a path to circumvent the least cost requirements within their policy authority. Likewise, some legislative bodies have passed legislation that requires a specified percentage reduction in consumption to be captured by energy efficiency. Yet few, if any of these legislative initiatives have provided policymakers with the authority to move from a legislative objective to a portfolio approval mechanism that allows the calculated value of energy efficiency to be an adjustable metric in considering how a jurisdiction will meet those objectives. Policymakers report that they are given (or give themselves) a charge to reach an objective without the policy tools to accomplish the task. Or, when they are given the tools to do so, they report that their careers are placed at risk if they too aggressively move to reach those objectives. They also report substantial pressures from the fossil fuel industry not to go too fast with energy efficiency or to allow program costs to move above the price of carbon based generation, the very fuel that efficiency is trying to avoid under a climate initiative.

We are not saying policymakers are misguided or dishonest; quite the opposite. But it is clear that policy priorities and evaluation rules can be, and frequently are, in direct conflict.

Program Design & Evaluation Wish List

So what should policymakers do to resolve this conundrum? We offer some proposals, based on best practices in evaluation and among the jurisdictions with which we have worked.

Align Short Term Policies with Long-Term Goals

While this is easy to say, it is difficult to do. Policymakers, as we've pointed out, are subject to an array of competing pressures, and have multiple constituencies to whom they answer. They exercise leadership and vision in setting long-term goals that address climate change, load growth, consumer protection, etc. But they also make decisions on the rules of the game, where it is usually acceptable to err on the side of conservatism. The problems arise when these rules conflict with the long term goals. Either the long term goals become unattainable, or the short term policies to set the evaluation rules need to be smarter and, perhaps, less conservative. Letting these mismatches remain can be self-defeating for the policymakers (not to mention crazy-making for program managers and evaluators). We cannot recommend all the adjustments that should be considered, and every jurisdiction is different, but we have pointed out some of the consequences of on-going mismatches between the short-term policies and the long-term goals, and these suggest some adjustments that could be adopted.

Match Program Theories to Evaluation Methods (and vice versa) -

Assuming the policies and goals can be brought into better alignment; there remain a set of more detailed problems to solve between implementers and evaluators. These are not staff-level problems, but management problems for the portfolio managers and the evaluation managers to resolve. As we've asserted above, portfolio managers will design their programs to, among other things, reduce evaluation risk and maximize the likelihood that they will be credited with the maximum amount of net savings allowed by the rules. Evaluators, by and large, are sympathetic to the goals of energy efficiency, but are charged with applying rigorous evaluation methods to determining the outcomes of the implementers' efforts. In doing this, they apply evaluation protocols, which reflect the evaluator's professional practices within the limits of the policy choices made by the policymakers. Both implementers and evaluators are interested in aligning the program goals and the evaluations, so that the savings are fairly and accurately credited. When there are mismatches, e.g. when a program goes after long-term savings but the evaluators' protocols say that those long-term influences produce free-riders, then neither the program nor the evaluation are entirely successful.

Best practices in energy efficiency suggest that program designers carefully explicate their program theories, describing the market barriers they seek to overcome and identifying specific testable hypotheses that the evaluators can examine to determine program success. Program implementers, in turn, should be documenting and collecting data to demonstrate how they are addressing those barriers, and to show evaluators how they have been successful. Evaluators apply time-tested evaluation practices (sampling, survey research, engineering reviews, field studies, etc.), as well as their best professional judgment, to test whether, and to what degree, the program implementers have succeeded in executing their program theories and capturing their savings. Best practices suggest that all three levels of this process be engaged in an early and on-going collaboration to ensure that what is done and what is evaluated match up.

The all-too-common practice of designing and implementing a program for several years, and then calling in an evaluator to figure out what was done and what can be measured to estimate savings allows for too many mismatches, and can produce less-than-useful evaluation results. The consequences of inaccurate or inappropriate evaluations can call the entire energy efficiency enterprise into question.

Drop Adversarial Stance in Favor of Collaborative One

An underlying theme that runs through all of the examples discussed in the paper is that there is often an adversarial relationship among policymakers, portfolio managers and evaluators. The reasons for this are many and diverse, but unnecessary. Often, it is an institutional problem. The classic example of this is when the policymakers are public utility policy agencies, and the portfolio managers are investor-owned utilities. In some jurisdictions or for some companies there is a long and rigid history of adversarial relationships; that's how they deal with each other. There is institutional distrust, there is an expectation that the regulatee will do exactly as the policymaker dictates (if all you have is a hammer, every problem looks like a nail), and this can get in the way of implementing programs that will successfully interact with the energy efficiency market and the building industry. But this is not the only example. There are cases where policymakers and implementers have become adversarial simply from a history of conflict between the participants that stems from the need to make a profit for stockholders and the need to control costs for ratepayers, both while providing reliable energy resources.

There are also positive examples in many states, where there has been an overriding atmosphere of collaboration toward the common goal of energy efficiency, and a history of positive working relationships. In those cases, evaluation is a tool used by the policymakers to help the implementers do a better job, by providing early feedback that leads to actionable recommendations. In jurisdictions where there is substantial friction between the policymakers and the utilities, the evaluation is seen as a way to gain some level of additional control over an adversarial relationship. In these cases evaluation accuracy, objectivity and reliability can take a back seat to other priorities.

Establish an Evaluation Framework That Is Helpful to the Portfolio

In discussing problems with the traditional evaluation framework, and with the policies that guide the protocols operating within that framework, it is important to remember that nothing in this paper is meant to suggest that program evaluation is not necessary or required, or that evaluation only needs to be conducted at the portfolio level or at the broader market level. As noted in Mr. Hall's IEPEC paper [Hall], portfolios are built on programs and programs provide the direct interface between a program's service offering and the customer's acceptance of that offering. Managers design, implement and adjust programs within a portfolio. That said, programs should be defined broadly enough to include related activities that affect the same market sectors, rather than fragmenting them into smaller pieces that are difficult to evaluate separately. The evaluation effort has to provide programs with the information they need to be well managed and cost effective, as well as to document the impact of the program and the portfolio. Both the programs' impacts and the portfolio's impacts need to be assessed if the evaluation effort is going to document total net energy impacts. The portfolio's impacts are more than the summation of the direct impacts of a portfolio's resource programs, but rather are the total impacts induced by the portfolio's efforts within a set of market effects caused by those initiatives. Evaluation timing is also important. When the evaluation effort is delayed so that the findings are provided years after the program decisions are made, the ability of the portfolio to be cost effective and achieve its energy and carbon goals is eroded.

In establishing an evaluation framework, it is important for policymakers to understand that the evaluation effort is a part of effective program design and operations, and so is a key part of the process for building effective portfolios. It is not enough to effectively design and manage the evaluation studies; the timing of the presentation of the evaluation findings is critical to successful programs and portfolios. An effective evaluation framework must be embedded within policies that enhance the ability of the jurisdiction to achieve its energy and carbon reduction objectives. Focusing the evaluation framework on a single metric, such as calculating an incentive payment or reaching an arbitrary savings goal, rather than looking for information to improve the programs in that portfolio, can do more to harm that portfolio than poorly designed and operated programs. The policies on which evaluation frameworks are based need to avoid a myopic perspective, but look instead to the broader needs for reliable program, portfolio and market evaluation findings, so that the true effects of those efforts are recognized. Policymakers should ask for net portfolio savings, rather than focusing the attention of their evaluation contractors solely on the impacts of the individual programs operating within the larger portfolio.

The opportunities to achieve greater energy savings, and to substantially reduce carbon emissions, are challenging our evaluation frameworks, but it is unclear that traditional evaluation frameworks are ready for these opportunities.

References

[ACEEE] Kushler, Martin, D. York & P. Witte. 2009. **Meeting Aggressive New State Goals for Utility-Sector Energy Efficiency: Examining Key Factors Associated with High Savings**. American Council for an Energy Efficient Economy, ACEEE Report No. U091.

[CPUC] California Public Utilities Commission. **2004-2008. National Energy Efficiency Best Practices Study**. http: www.eebestpractices.com

[Hall] Nick, Hall, Patrick McCarthy. 2009. **Portfolio Evaluation Versus Program Evaluation: Is There a Balance?** International Energy Program Evaluation Conference, Conference Proceedings.