

Emerging Issues in the Evaluation of Energy Efficiency Programs: The US Experience

Edward Vine, Lawrence Berkeley National Laboratory and California Institute for Energy and Environment

Nick Hall, TecMarket Works

Kenneth M. Keating, Independent Consultant

Martin Kushler, American Council for an Energy-Efficient Economy

Ralph Prah, Panoptic Energy Consulting

ABSTRACT

The evaluation, measurement and verification (EM&V) of energy efficiency programs has a rich and extensive history in the United States, dating back to the late 1970s. During this time, many different kinds of EM&V issues have been addressed: technical (primarily focusing on EM&V methods and protocols), policy (primarily focusing on how EM&V results will be used by energy efficiency program managers and policymakers), and infrastructure (primarily focusing on the development of EM&V professionals and an EM&V workforce). We address the issues that are currently important and/or are expected to become more critical in the coming years. We expect these issues will also be relevant for a non-US audience, particularly as more attention is paid to the reliability of energy savings and carbon emissions reductions from energy efficiency programs.

Introduction

The evaluation, measurement and verification (EM&V) of energy efficiency programs has a rich and extensive history in the United States, dating back to the late 1970s. Most of the energy efficiency programs and related evaluation in the late 1970s and early 1980s occurred at the federal and state government level (e.g., US Department of Energy's Weatherization and State Energy Programs). Since then, most efforts occurred at the state level in the context of energy efficiency programs funded through utility rates. (In the U.S., each individual state regulates the utilities that operate within its borders.) In the last few years, with the arrival of federal "economic stimulus" funds, the government has become much more active in promoting energy efficiency. For governmental energy efficiency programs, the principal driver promoting evaluation was legislative: ensuring that energy and environmental goals (e.g., reducing GHG emissions) were met. For utility energy efficiency programs, the principal driver promoting evaluation was regulatory: ensuring that utilities meet their targeted goals cost-effectively and efficiently.

During this time, many different kinds of EM&V issues have been addressed, with a primary focus in the U.S. at the program level (rather than at a policy level): technical (primarily focusing on EM&V methods and protocols), policy (primarily focusing on how EM&V results will be used by energy efficiency program managers and policymakers), and infrastructure (primarily focusing on the development of EM&V professionals and an EM&V workforce). We address the issues that are currently important and/or are expected to become more critical in the coming years¹. We expect

¹ Due to page limitations, we could not focus on other important evaluation issues, such as closing the loop between evaluators and implementers, the evaluation of persistence, whether there is a rebound (takeback), the evaluation of

these issues will also be relevant for a non-US audience, particularly as more attention is paid to the reliability of energy savings and carbon emissions reductions from energy efficiency programs. While the EM&V profession has developed and matured over the years, many of the EM&V issues described in this paper have not been resolved and, in some cases, are highly contested – in and outside of regulatory proceedings. In fact, the authors of this paper do not entirely agree on how these issues should be settled, despite their extensive experience in this field. Evaluation is a risk management enterprise, and it is important to remember that there is a lot of uncertainty in what evaluators are asked to address. It is the job of the evaluator to use sound EM&V techniques to reduce uncertainty to the extent possible, and provide reliable and practical information for decision-makers.

1. EM&V Technical Issues

Based on the experience in the U.S, the most important and transferable technical issues are “net savings” (incrementality), evaluation of market transformation programs, and evaluation of the carbon impacts of energy efficiency programs. Of these, the most developed and the largest technical issue is the evaluation of net energy savings (versus gross energy savings). We provide an overview of the key issues involved in calculating net energy savings which take into account direct energy savings but also energy savings from free riders and from spillover – controversial issues dealing with what is included in net energy savings and when they should be used (see Titus and Michals 2008). The second technical issue deals with the evaluation of market transformation programs (versus resource acquisition programs), focusing on education, information, training, and leveraging collaboration with manufacturers, as well as incentives. We review the key data collection and analytical challenges in evaluating market transformation programs, including sustainability. The third technical issue deals with the conversion of energy savings to the reduction of carbon emissions. We compare the different approaches for calculating carbon emission savings on the basis of accuracy, costs, and complexity.

1.1. Net energy savings calculation

The concept of net energy savings is fairly simple: what were the true effects produced by a program or intervention in terms of energy savings, separated out from what would have otherwise occurred absent the program or intervention? Unfortunately, this simple concept is exceptionally difficult to measure in practice, particularly in a way that meets specific reliability standards for accuracy or comparability. This problem is compounded in current practice in the U.S. because there are two general conditions that impact the ability to estimate net impacts. These are the questions of definition and technical measurement.

The definition of what constitutes net energy impacts can be state-specific, in some cases program-specific, requiring the measurement approach to be tailored to meet the applicable definition for a specific regulatory jurisdiction. The difference in definitions can have a substantial impact on the estimate, as well as on the evaluation method that is used. For example, in California (2004-2009), net energy savings are defined by the California Public Utilities Commission to be gross energy savings minus the energy savings from free riders. In this case, the gross energy savings are reduced to account for what a specific evaluation methodology can identify as a program-induced installation, subtracting out savings from installations that are driven by other factors. The following formula represents the current California definition:

$$\text{Net savings} = \text{gross savings} - \text{free riders}$$

On the other hand, in New York, net energy savings are defined by the New York Public Service Commission as gross energy savings, minus savings from free riders, plus energy savings due to participant spillover and market effects. Participant spillover is the savings from program participants who, as a result of the program, installed additional energy efficiency measures, but who did not obtain a program incentive for those additional measures. Market effects are the market level savings that resulted from program influences on the market and the operations of that market (sometimes referred to as nonparticipant spillover, since these end users did not participate in the program and did not obtain a program incentive for those measures), but the market for energy efficiency was affected by the program. The following formula represents the New York definition:

$$\text{Net savings} = \text{gross savings} - \text{free riders} + \text{participant spillover} + \text{market effects}.$$

In some states, market effects are not equivalent to nonparticipant spillover, since program participants as well as nonparticipants are affected by market effects. For example, in Wisconsin, depending on the program, the evaluation of net savings may focus either on: (1) free riders only, (2) free riders and participant spillover only, (3) free riders, participant spillover, and nonparticipant spillover, or (4) total market-level net impacts, without any effort to disaggregate by spillover type.

Because the market effects of a program can be as large as or larger than the program's gross savings, the resulting quantification of net effects from one state to another can be very different for the same program, rebating the same measures, targeting the same customers. The definitional difference alone makes comparing a net effect from one program to the next problematic, particularly if the evaluation approach varies from state to state. Similarly, in a carbon-focused world (see Section 1.3), the definition of net effects can result in large and significant differences in reported carbon reductions resulting from the same program operating in two different jurisdictions. Once the definitional issue is addressed, typically through a regulatory decision establishing the definition of net savings for a specific state, the technical issues associated with measurement must be addressed. The measurement of net energy savings can be accomplished using a variety of different approaches. However, one of the most important concepts to understand within a technical measurement approach is that net savings is a behavior metric that adjusts gross savings to account for how a program influences the decision-making processes of the participants or people in the marketplace. Thus, net savings evaluation approaches measure changes in decision behavior, and the evaluation approach must document how the program changed end users' decision behaviors.

One final aspect of this subject that is receiving increased discussion within the U.S. is the issue of whether the concept of net energy savings itself is becoming less important. To illustrate this issue, consider the situation of a specific utility energy efficiency program. Under the historical regulatory paradigm where energy program evaluation grew and thrived in the U.S., the "frame" that was established was the need to carefully measure the net impacts of the program in order to quantify its delivered energy savings. This was important both to determine the remaining supply resources that the utility system may need to acquire as well as to demonstrate that the energy efficiency program was "cost-effective." (The latter function tended to receive disproportionate focus, in large part because utilities and certain other key parties tended to oppose requirements to include energy efficiency programs – so energy efficiency was subjected to a burden of proof that other resources generally did not face.)

The current context for energy efficiency programs in the U.S. is evolving into something much different. First, energy efficiency has proven itself as a cost-effective resource, and is widely regarded as the least-cost utility system resource available; so much of the historical threshold 'burden of proof' intensity of scrutiny has faded in many (but not all) states. Second, the overall policy objectives for energy efficiency are broader now – in particular, in the area of climate change. At the end of the day, what is important is that atmospheric carbon loading is reduced below some target level. For some, it is much less critical to parse out who is responsible for precisely what share of that reduction (however, for others, this parsing out is still important in ensuring that programs are effective in reducing emissions). And third, the mosaic of public policies and market interventions directed at achieving energy efficiency now are complex, it may be impossible to sort

out the net effects of a single program operated within a specific program funding cycle. Consider a typical situation in the U.S. today. It is not uncommon to simultaneously have: public awareness energy efficiency messaging by the state government; federal tax credits for energy efficiency measures; private-sector messaging and financial incentives for energy efficiency products; utility audit and informational programs; utility rebate and incentive programs; and general media coverage of energy efficiency related issues. In this context, some are beginning to believe that it is a “fool’s errand” to try to isolate out the effects of any one policy or program on an individual’s behavior.

This is not to say that there is no role for estimating net impacts. The evaluation of net program effects is important for improving the effectiveness of programs. That is, while the energy impact results of net analysis may not be accurate across the various types of change-inducing efforts, the results can be used to assess the effectiveness of the program in targeting non-freeriders compared to other programs. Similarly, net effects will still be important in any circumstance where specific financial remuneration is tied to specific amounts of energy savings. But for overall policy objectives, the tide is turning toward an emphasis on maximizing the aggregate gross impact of all policies and interventions. In conclusion, the most likely direction in the near-term will vary by state: some states will focus on gross savings while others will continue to rely on net savings. It will then be left to the policymakers in each state to select the type of savings to report as well as the amount of evaluation resources that will be devoted to each approach. It remains to be seen whether a dual approach will be viable at a national level, if national initiatives (e.g., energy efficiency resource standards, cap and trade for reducing GHG emissions, or a national EM&V protocol (Section 2.3) are achieved.

1.2. Market transformation evaluation

The concept of precisely identifying net savings becomes even more challenging when the goal of energy efficiency policies or programs is to change the market. Strategic market transformation entails a deliberate effort to take advantage of leveraging points in the market structure and partnering with other market actors to create a large scale and diverse impact on market structures and choices. While the strategy may be planned and initiated centrally, the moment it moves into the market a market transformation effort, by its nature, depends on a lot of actors, motives, and interplay among competing interests. The EM&V of market transformation initiatives usually doesn’t attempt to isolate and apportion a specific level of savings to each of the unknown variety of market actors, but entails making sure that the goal, or progress toward the goal, is achieved and that the ratepayer initiative was an important part of achieving the goal. It should also make sure that savings from resource acquisition programs aimed at the same targets are not double-counted for state or regional accomplishments.

Market characterization and assessment are key evaluation activities. In characterization, we need more work on describing specific markets or market segments, including the types and number of buyers and sellers in the market, key actors, type and number of purchases and transactions that occur. In assessment, we need more work on examining trends in the market over time including (1) changes in the structure or functioning of a market and (2) the behavior of participants in a market that results from one or more program efforts. Critical to this assessment is the development of market theory and program theory: e.g., logic models have been developed to show how energy efficiency programs and markets operate and are linked by looking at inputs, outputs and intermediate variables. Thus, once these linkages have been identified, market indicators can be developed that relate the adoption of energy efficiency products, services or practices to program activities. Data collection on these indicators is expensive and challenging, but valuable (see Vine et al. 2009 for a recent analysis of market effects).

Sustainability is an integral element of market transformation – changes in market structure and operations, and how the changed market contains mechanisms to sustain the market effects. There needs to be a consideration of specific changes to the market that will help “lock-in” savings

(Hewitt 2000). And charting a path to a transformed market will require specific planning, implementation, and evaluation activities.

1.3. Carbon emissions calculation

Based on the experience in the U.S., there are four generally accepted approaches for documenting the amount of carbon emissions saved as a result of energy efficiency programs:

1. Average carbon multiplier approach (carbon emissions factor)
2. Hourly weighted average carbon multiplier approach
3. Hourly dispatch carbon emissions calculation approach
4. Oxidation reduction equation approach (heat-rate approach)

The first three approaches apply to electric savings, while the fourth applies to non-electric savings. The average carbon multiplier approach assumes all kWhs are saved in a way that equally reduces carbon emissions, no matter when the savings occur. In this approach, an average carbon multiplier is estimated based on the average fuel source used to generate the average kWh of electricity over the projected life of the predicted savings. The total kWh savings are then multiplied by the average carbon reduction within a jurisdiction of generators, typically a utility service territory.

The hourly weighted average carbon multiplier approach is similar to the non-weighted approach above, however, in this approach, the average carbon reduction per kWh is calculated for each hour of the day over the expected life of the savings and then multiplied by the kWh savings for each of those hours. This approach requires a carbon emission multiplier estimated for the expected generation mix for each hour of the year over the life of the expected savings. The savings are estimated for each hour of the year, typically for 15 to 25 years, so that each hour can be multiplied by the carbon reduction multiplier for each hour. To simplify these two approaches, it is assumed that the impacts that are estimated for the first year of savings also apply to all other years in which the savings occur.

The hourly dispatch carbon emissions calculation approach uses generator-specific dispatch data provided by the dispatch operators located at generators providing the power, so that the carbon savings are based on the generators that are typically on-line and operating for each hour of the year. If the multipliers used for the second approach (above) are based on dispatch data for each hour, then approaches 2 and 3 are almost identical in their results. However, for approaches 2 and 3 to be used, the evaluation must provide hourly savings load shapes over the effective useful life of the installed actions as an evaluation output, rather than projecting only annual savings or effective useful life savings.

All of these approaches produce estimates based on an assumed generation mix for the life of the expected savings. For each of these three approaches, the fuel saved and the carbon reduction estimated is based on the amount of fuel used to generate the electric power. For example, consider the generation of electric energy that is about 25 to 35 percent efficient, requiring about three to four times the resource energy than the amount of energy contained in the kWh consumed. If an energy efficiency program saves 100,000 kWh per facility, then 300,000 – 400,000 kWh (or 1,024 to 1,365 MBtu) does not have to be consumed at the power plant. Because of the inefficiencies of electric generation, savings of electricity that is generated from coal, oil or natural gas fired generators can save large amounts of carbon.

The oxidation reduction equation approach is used for non-electric savings and can apply to gasoline, fuel oil, propane, natural gas and other combustible fuels. This estimate multiplies the units of energy saved (e.g., gallons of gasoline) by the amount of carbon released by burning that fuel. This approach assumes that all of the carbon in the oxidized fuel is converted to carbon dioxide either in the combustion process or when unoxidized carbon is emitted to the atmosphere and is subsequently converted to carbon dioxide.

There are several uncertainties associated with these approaches that can affect the accuracy

of the savings estimates. The key uncertainties are associated with not accurately:

1. Estimating energy impacts
2. Knowing what fuel type is saved
3. Knowing the efficiencies of the generation facilities impacted
4. Knowing the hour the savings occur over 8,760 hours a year
5. Knowing the generation mix for any given hour over the effective useful life of the savings
6. Knowing how generation facilities are cycled or how to accurately predict unit cycling and the relationship to demand and savings
7. Knowing the effective useful life of the savings projections
8. Knowing the use or dispatching of renewable energy supplies over the effective useful life of the expected savings

For these reasons, estimates of carbon savings should generally be considered proxy estimates of actual impacts rather than documentation of achieved impacts. Also, because of these uncertainties, and the resulting possible estimation errors, it is often considered prudent to estimate carbon impacts using the least expensive approach for the accuracy desired. For example, the average carbon multiplier approach is considered to be the least expensive approach as well as the least accurate approach, while the hourly based approaches are relatively more expensive and more accurate. Furthermore, because most evaluations conducted in the United States now provide only annual or effective useful life savings rather than load shape savings, most of the current evaluation efforts will rely on the average carbon multiplier approach rather than on hourly or dispatched based approaches. Evaluations that provide load shape reductions for each hour of expected savings can move to the hourly based approaches. However, using hourly approaches that are based on dispatch data that are one to two years old, may mean that the cost of using dispatch data is not worth the added reliability benefits, especially during periods of rapidly changing generation mix or when that mix is expected to change. Recent work in Wisconsin (Rambo and Ward 2009) concluded that the value of matching energy savings with emissions on an hourly basis is probably not worth the effort if load shapes must be developed specifically for the avoided emissions estimate. On the other hand, if the load shapes are already developed as an important input to assigning avoided costs to energy savings in the analysis of cost-effectiveness, then it is relatively easy to apply the load shapes to avoided emissions as well. This is an area that U.S. evaluators could learn from the experience of other countries.

2. EM&V Policy Issues

The first policy issue deals with how energy efficiency programs are evaluated from a public policy perspective. We review the different metrics that are used by policymakers for evaluating energy efficiency, focusing on whether the total resource cost test is the best metric for addressing the challenges ahead (such as mitigating the effects of climate change). The second policy issue examines how the practice of evaluation depends on how the results are used: for example, (a) demonstrating energy efficiency as a reliable energy resource, (b) using energy efficiency as a means for reducing carbon emissions, (c) determining shareholder incentives, (d) improving the quality of programs, etc. The third policy issue deals with the challenges associated with the consideration of national EM&V protocols – and the effect that a national cap and trade policy might have on this issue.

2.1. Evaluation metrics

Metrics here are used as measures of success or progress toward goals. Metrics for acquisition programs may include gross savings (reductions in energy use that are related to the implementation of efficiency measures); net savings (savings that would not have occurred in the

absence of a program intervention); cost-effectiveness as measured by the cost of a resource compared to other energy resources (Total Resource Cost); the cost of the resource to the implementing agency (program administrator or utility cost test); or a societal test based on all benefits and costs that can be assigned values. Market transformation programs use metrics that look at progress toward changing a market, improving market share, and the sustainability of the market changes that results from program intervention.

In the evolving world of energy and environmental policy, we shouldn't be surprised if there are shifts in emphasis that question program goals, their metrics, and how evaluations are conducted. Historically, in many parts of the US, the emphasis has been on (1) efficiency paradigm, (2) programs, (3) net savings, and (4) the Total Resource cost (TRC) test of cost-effectiveness². With the growing importance of GHG reduction, each of these may have to be re-considered, as discussed below.

First, a paradigm based on energy efficiency (reduction in consumption relative to what it was or what it might be) does not align well with the desire of an absolute reduction in GHG emissions or policies aimed at “net-zero” consumption. Retrofitting the worst wasters of energy surely helps, but making a massive home more efficient than it might otherwise be still results in an increase in consumption and GHG emissions. If policy makers set the goals to be an absolute reduction in consumption and emissions in order to reach a specific target, then the practice of evaluation should try to measure absolute consumption. Evaluation follows the metrics.

Secondly, the emphasis on “programs” is too restrictive for policies that hope to operate in the consciousness of consumers, in the marketplace, and to create dramatic change in consumption through synergism. Even the concept of market transformation programs assumes that there is a directed approach that is limited in scope. By expecting all or most of the change to come from the limited implementation of programs narrows the practice of evaluation. Instead, the focus should be on the marketplace, and the evaluator needs to evaluate how the market is changing over time with respect to energy efficiency: in buildings, retail, manufacturing as well as in the educational fields and in the training of professionals (architects and engineers, finance and insurance industry, etc.).

Thirdly, as noted earlier, net savings is a concept of attribution that is both very hard to measure in a complex world of confounding influences (outstripping the capability of social science to provide reliable answers), and one that policy makers with a view toward the end-point of reduced emissions may find less relevant in the future. However, the concept of examining attribution is always important for making sure that the ratepayer/ taxpayer/societal resources are being spent prudently as well as making sure that resources are being used as effectively as possible on programs that reduce emissions. If one were to move away from net savings in energy policy, this would run counter to developing environmental policies that reward entities for “incremental or additional” GHG reductions – which is one of the key concepts underlying the Clean Development Mechanism in the Kyoto Protocol. But, as noted previously, it is time to revisit net savings – not only in the energy efficiency arena but also in the international discussions on climate change. If urgent and comprehensive efforts are needed, and the efforts by everyone need to be encouraged and noted (including free riders), then it may be necessary to live with the “extra costs” of paying people to reduce their energy use and emissions, even if they were already planning to do so. The other option is to be “smarter” by strategically making decisions upfront on what should be funded and what should not be funded (e.g., promoting energy efficiency measures and programs that will reduce the probability of free riders) and reduce the evaluation resources used to test for additionality for those programs over time.

Lastly, the concept of a Total Resource Cost (TRC) test, whether of programs or resources, is a regulatory paradigm that is designed to making sure that ratepayers/taxpayers are receiving the

² There is a long-standing difference in practice between parts of the country, like California, that apply TRC to programs versus those, like the Pacific Northwest Power Planning and Conservation Council, who apply it to the efficiency resource regardless of who pays the cost. The corollary is that a program-centric approach focuses on net savings, and a resource planning perspective focuses on the gross savings.

least cost resource, not necessarily the resource of best choice for reasons outside of regulatory purview. There are concerns with the continued use of the TRC – some argue that a new cost-effectiveness metric be used or that the TRC be changed by making significant changes to the inputs: e.g., avoided cost, discount rate, value of carbon emissions, measure lifetime (Hall et al. 2009), and non-energy benefits. For example, until recently, the benefits of the TRC were based on the avoided cost of a new generation plant – in California, that plant was the combined cycle gas turbine (CCGT) while in other states it may be coal-fired generation³. However, if climate change is the focus of national policy and we are to reduce our carbon emissions, then some suggest that a renewable energy plant should form the basis for avoided cost calculations – which would make energy efficiency more attractive, since most renewable energy supplies are more costly than carbon-based generation. Furthermore, in some states, carbon adders are used in cost-effectiveness tests to try to account for the negative impacts of carbon-based fuels (coal, oil). However, the real cost of climate change is undoubtedly larger than the costs reflected in the carbon adders.

Another concern that has more recently arisen regarding the TRC test is that it has tended to be applied in an asymmetrical fashion. That is, customer direct costs for the energy efficiency measure are virtually always added in to the total cost side of the TRC equation, but the additional customer benefits beyond utility system resource savings (e.g., increased productivity, reduced maintenance costs; aesthetics; etc.) are very seldom quantified and added in to the TRC calculation (in part because they are more difficult and expensive to measure). This has led some to call for the use of a “utility cost test” (aka “administrator’s cost test”), which is more comparable to the way that other utility resources are judged (i.e., costs and benefits to the utility system).

The purpose of discounting is to bring all costs and returns at different points in time to a net present value, so that different investment choices with different costs and returns can be compared. The discounting values typically reflect the perspectives of the key stakeholders involved in managing risk (e.g., the utility perspective versus the public agency (societal) perspective). Some argue that the current approach for discounting the value of future savings supports the analysis of short-term economic decisions but does not support the analysis of long-term decisions like climate change where the impacts occur over a much longer time period. Very small (negative?) discount rates have been proposed as one solution, making energy efficiency more financially attractive.

As noted above, some states have used carbon values in their benefit cost tests, but the values are based on a traded value of carbon, a proxy to represent an expected trade value (e.g., within a cap-and-trade system), or a derivative of a traded or expected trade value. Although we do not know the real value of avoided carbon emission, the value may be significantly higher if all of the environmental costs (or the costs of achieving a sustainable environment according to some analysts) are included (instead of \$3 to \$45 as reflected in carbon markets in Europe and in the Northeast and which exclude all of the environmental costs), making energy efficiency an even more viable solution for addressing climate change.

The effective useful life (EUL) of a measure (i.e., measure lifetime) is the period of time that the measure is expected to perform its intended function in a typical installation. Some estimates of EULs are conservative and underestimate the actual lifetimes of measures. As a result, measures that are very effective and long-lived (e.g., windows, insulation and new building envelopes) are not recognized or valued as highly as they should: while their initial costs are reflected in the benefit-cost ratio, their long-term savings are reduced (e.g., 20-25 years, instead of 40-50 years). On the other hand, estimates for some measures may be too optimistic, not accounting for business turnover (which may cause measures to be removed prematurely) or simply relying on manufacturers’ estimates which have not been supported by field observations. Careful analysis of measure lifetimes is warranted to ensure that there is no bias overall.

Non-energy benefits (or costs) – such as reduced emissions (see above) and environmental

³ In addition to the forecast of capital cost of facilities, natural gas prices over 20 years are the most significant component of avoided costs. California is now using market gas prices for the near term as the avoided cost for both gas and gas-fired combustion plants.

benefits, productivity improvements, high comfort and convenience, reduced debt and lower levels of arrearage, and job creation - are typically not included in benefit-cost tests. Some argue that these non-energy benefits should be included, since evaluation methods are available and, more importantly, these benefits often are valued more highly than the energy benefits for motivating end users to invest in energy efficiency or change their energy behavior (Skumatz et al. 2009).

In conclusion, as the scope of energy-related policy and investments both expand and intersect with energy and environmental regulation, they may force cost-effectiveness metrics to be re-defined, with implications for EM&V.

2.2. Evaluation practice

The practice of evaluation depends on how the results are used, as noted above: for example, (a) demonstrating energy efficiency as a reliable energy resource, (b) using energy efficiency as a means for reducing carbon emissions, (c) determining shareholder incentives, (d) improving the quality of programs, etc. The number and types of stakeholders have increased over time, making the evaluation practice more comprehensive and of greater interest to parties who wish to use the evaluation results for their own agendas. For example, if certain stakeholders either do not value energy efficiency as a resource or question the cost and/or reliability of energy efficiency as a resource, then it is incumbent upon the evaluator to accurately quantify the costs and benefits of the energy efficiency program and to compare the value of that resource (e.g., the cost of conserved energy) with other resources (e.g., the avoided cost of a combined cycle natural gas plant, a wind generator, or some other generation source). In addition, evaluators need to periodically assess the persistence of these energy savings over time, to ensure that energy efficiency can be counted in utility procurement plans and to make sure that the carbon reductions are properly accounted for over the life of an energy efficiency measure. Similarly, if stakeholders are primarily interested in shareholder incentives, then the evaluation will focus on those measures that affect the ultimate outcome of the incentive mechanism (e.g., high impact measures), and may pay little attention to those measures (and programs) that do not result in significant energy savings.

If stakeholders are interested in energy efficiency as a strategy for reducing carbon emissions, then the evaluator must be able to convert energy savings into carbon emissions (as described in Section 1.3). In addition, the appropriate cost-effectiveness calculation should be used, as determined by the policy makers: e.g., a revised Total Resource Cost test (see Section 2.1). In fact, it may turn out that policymakers are interested in all energy efficiency programs (no matter the cost effectiveness) if they believe that climate change is a problem that needs to be addressed urgently and comprehensively.

Finally, if stakeholders are primarily interested in improving the quality of the programs, then more evaluation work will need to go into process evaluations (as well as impact evaluations) to ensure that the programs are delivering the energy efficiency measures efficiently and effectively. We need more research on which consumers participate or do not participate in energy efficiency programs and why. We need more research on behavior of the key stakeholders: how they use energy and how they make decisions on investing in energy efficiency. And we need more research on the overall market for energy efficiency products and services – how is it changing and how have programs affected the market. Historically, process evaluations and market assessments have been perceived as less important than impact evaluations, especially if the attention is devoted primarily to energy savings impacts.

2.3. National EM&V protocols

Although EM&V guidelines have been prepared by several organizations at the state, national and international levels (e.g., TecMarket Works 2004; CPUC 2006; NAPEE 2007; USDOE 2007), there are no national EM&V protocols that organizations are required to follow. Given the

interest in energy efficiency resource standards and potentially a national cap and trade policy/program and the need for ensuring high quality standards for conducting evaluations, there is renewed interest in having a national EM&V protocol. The elements of the national protocol would include: common evaluation terms and definitions; common evaluation methods; common savings values and assumptions (e.g., energy, costs, measure life, and persistence); guidelines in savings precision and accuracy; and common reporting formats.

The key strength of national evaluation protocols is that they allow energy savings results to be grounded within an assessment approach that can produce reliable and transparent savings estimates if the protocol is based on rigorous evaluation practices. National protocols (using a consistent set of inputs and reporting formats) can also allow savings to be compared from one state to another or from one evaluation to another. If the research is based on the same protocols, in theory, the results should be comparable – assuming that the protocols are detailed enough to prescribe the required evaluation approach. Similar approaches also reduce evaluation estimation error risks (increasing the credibility of energy efficiency) and reduce evaluation costs to states that wish to use these approaches. A national protocol would also minimize confusion for and reduce barriers for the growing market of energy efficiency providers (the transaction costs are high for providers who have to meet different state-mandated evaluation requirements and levels of rigor).

Comparability and compatibility assume that the definitions for what constitutes an achieved impact are also identical. Because many states define net energy savings differently (see Section 1.1), a protocol that prescribes a reliable evaluation approach, but is applied to different definitions of net energy impacts, will not provide comparable results. It is not enough to prescribe an evaluation approach to achieve comparability and compatibility, the definitions on which that protocol is based must also be prescribed. Finally, similar approaches also reduce evaluation estimation error risks and reduce evaluation costs to states who wish to use these approaches.

In establishing a protocol, it is also important to place into that protocol a prescriptive approach for dealing with the conditions that most impact the reliability of the findings. It is important to remember that energy efficient measures do not provide savings on their own. Savings are produced only after measures are integrated within a customer decision and operational environment that provides savings to be measured. The authors of this paper have seen identical programs, implemented in similar climates, and serving similar customers result in substantially different energy impacts. If the policy makers want to understand why programs produce the savings they achieve, then the protocols must also focus on prescribing evaluation approaches that support this purpose.

There are other concerns that may prevent the adoption of a national protocol if one were to be developed and implemented. First, engaging a broad range of stakeholders is challenging: standards need to be perceived as fair by a diverse set of stakeholders, although some may see the protocols as restricting states from doing what they want. Moreover, it may be difficult to get consensus among all of the stakeholders on some of the key issues mentioned previously: e.g., rely on gross energy savings or net energy savings; if net energy savings are used, include free riders and spillover or just free riders; etc. Second, a national protocol may be seen as impeding innovation in evaluation practice at the state level or inadvertently exclude evaluation practices that are valid. Third, best achievable practices in evaluation may differ from one region to another, due to resource availability (see below). Similarly, a national protocol may be viewed as too stringent for some states, and too lenient for other states. Fourth, a national protocol may end up as too general and not specific, for the reasons above and because savings algorithms and assumptions will vary by program design (see below). And fifth, a national protocol may increase transaction costs if states need to respond to state reporting requirements and goals as well as national requirements.

To develop a national protocol, the national evaluation standards must be developed objectively by third parties and build in room for flexibility and opportunity for updates. The protocols must also ensure that state goals and reporting needs are being addressed. Thus, an open and transparent process with opportunities for stakeholder input and participation needs to be

encouraged and the crafters of the protocol must be willing to negotiate some elements (e.g., less rigor in the beginning versus more rigor in the future).

A key consideration in any move toward a national protocol must focus on the resources available to support that protocol's application. An evaluation protocol that is based on an evaluation budget that is eight percent of the portfolio's resources may not be applicable for evaluations of programs in states that have fewer evaluation resources. On the other hand, if the protocol is based on the lower ends of the evaluation budget spectrum (e.g., 1-3%), then states that need more reliable results must increase the evaluation rigor and move beyond the prescribed protocol (as well as be willing to pay for that reliability through an increased evaluation budget). In this case, the national EM&V protocol would provide an array of evaluation categories: set a minimum level of rigor for all programs, but encourage evaluations to go beyond the minimum level of rigor, if the desire and budget are available. This would be similar to California's EM&V protocols.

3. EM&V Infrastructural Issues

Based on the experience in the U.S., the most critical issue related to infrastructure deals with the development of a professional evaluation community and workforce that is able to address the technical and policy issues mentioned above. We briefly review the history of the International Energy Program Evaluation Conference (IEPEC), the most highly regarded conference in energy efficiency program evaluation, and demonstrate how this conference has supported the development of evaluation professionals and the evaluation community as a whole through the publication of peer-reviewed papers, training workshops, and the networking of evaluation experts with energy program managers and policymakers. The key challenge is how to train the next generation of evaluators. This is an area that U.S. evaluators could learn from the experience of other countries.

3.1. Developing a professional evaluation community and workforce

The International Energy Program Evaluation Conference (IEPEC) was organized in response to a need that was developing over several years, but which reached a turning point in 1983. Energy efficiency was a new field in the late 1970's in the US. Several programs were initiated by the US Congress in response to the oil crises of the 1970s. One was a requirement for all large utilities to provide informational audits to residential customers on how to save energy and money in their homes. Another federal program involved grants to states to improve the energy efficiency of institutional buildings. The evaluations of these programs were left up to the discretion of the states. When the federal government began a program to provide states with lump sums of money to run a variety of programs in 1982, they required that these programs be evaluated. However, several states (in particular, Illinois) and the U.S. Department of Energy (DOE) recognized that there was no infrastructure to provide the scale of evaluations required. In order to provide an opportunity for state grant recipients to receive training, the Illinois State Energy Conservation Program (SECP) and DOE provided money for an evaluation conference in Chicago for the summer of 1983. A group of individuals with a background and experience in evaluating energy and non-energy programs helped to organize the conference and solicit papers and presentations.

The success of the first conference and the enthusiasm of attendees led to the decision to organize follow-on conferences, at first annually, and later every two years. By 1988, the IEPEC was established as a non-profit, educational corporation. Over the years, more than 3,000 professionals with an interest in the evaluation of energy efficiency programs have attended the IEPEC conferences.

Over time, the conference organizers recognized the need for providing workshops and training that would inform new professionals, the managers of evaluation, and the users of evaluations. Topics ranging from introductory statistics, to planning and managing evaluations, to

measuring GHG emissions were offered on the day before the conference. Although the topics changed over time, the demand for the mini-classes never weakened. While other organizations have offered multi-day training and offered workshops ahead of their conferences (AESP and EPRI, for example), the IEPEC has remained the pre-eminent source of exposure to energy program evaluations.

The educational elements of the conferences go beyond formal workshops to include peer-sharing, refereed papers, poster sessions, expert panel discussions, and the all-important informal networking. The authors are confident that without the continuity and commitment of IEPEC, the professional development of energy program evaluators would have been difficult to achieve. It is this type of infrastructure that is needed in other countries to help to promote energy efficiency evaluation. The 2010 IEPEC is a good start in this direction, and we will watch to see how other countries develop their evaluation community and workforce.

3.2. Training the next generation of evaluators

In 2006, IEPEC and the Association of Energy Services Professionals (AESP) conducted an online survey of energy evaluation and market research professionals to characterize the energy evaluation and market research profession (Bensch et al. 2006). The evaluators noted that most of them learned their trade (evaluation) on the job – either they took an evaluation job (38%) or evaluation was a component of their non-evaluation job (29%). For others, evaluation was a topic in their academic field (9%) or they studied evaluation as an academic field (9%). On-the-job experience will remain critical for adding new people to the field of evaluation. However, with the increased activity in the energy efficiency arena and the need for trained evaluators, organizations are increasingly experiencing difficulty in finding people who are knowledgeable about and experienced in the evaluation of energy efficiency programs. Several organizations are starting to meet this need by providing evaluation training. The Efficiency Valuation Organization (EVO) offers a professional certification course on measurement and verification, as well as a course on the IPMVP (EVO 2010). The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) offers a measurement and verification training course (ASHRAE 2010). AESP offers a training course on the principles of research and evaluation (AESP 2010). Finally, IEPEC offers evaluation workshops one day prior to its evaluation conferences. As noted in Bensch et al. (2006), evaluators also rely on the evaluation literature and publications as well as attending conferences and workshops. As noted above, the premier evaluation conference is the one sponsored by the IEPEC; evaluation is also featured in other energy efficiency conferences and meetings held by the American Council for an Energy-Efficient Economy (www.aceee.org) and the Consortium for Energy Efficiency (www.cee1.org). In this manner, organizations can hire high potential employees who have been trained in energy efficiency or they can increase the skill levels of existing employees.

As noted above, universities and colleges can play a critical role in the education of future evaluators. Evaluators have diverse backgrounds – there is no one discipline that currently characterizes evaluators: they are drawn from many disciplines, such as engineering, architecture, sociology, geography, political science, environmental studies, ecology, economics, statistics, etc. In 2006, IEPEC prepared a directory of energy and energy-related programs at colleges and universities in the United States as a stepping stone for encouraging students' (high school, undergraduate, and graduate) involvement in the energy program evaluation field (www.iepec.org/IEPECHome.htm?links.htm). We hope that evaluators in other countries can develop their own directory to encourage students to pursue the field of evaluation.

4. Conclusions

This paper has examined key technical, policy, and infrastructure issues that are currently important

and/or are expected to become more critical in the U.S. in the coming years. While the focus of the discussion has been on programs and on the lessons learned in the U.S., we expect that many of these issues will also be relevant for a non-US audience, particularly as more attention is paid to the reliability of energy savings and carbon emissions reductions from energy efficiency programs. At the same time, we hope that U.S. evaluators will learn from the experiences of other countries as they delve into these issues.

References

- American Society of Heating, Refrigerating, and Air-Conditioning Engineers [ASHRAE]. 2010. “Determining Energy Savings from Performance Contracting Projects – Measurement and Verification” training course; website: www.ashrae.org.
- Association of Energy Services Professionals [AESP]. 2010. “Principles of Research and Evaluation” training course; website: www.aesp.org.
- Bensch, I., L. Skumatz, and E. Titus. 2006. “Evaluators Look in the Mirror: AESP/IEPEC Survey on Hot Topics and Industry Perspectives,” Presented at the 2006 ACEEE Summer Study on Buildings.
- California Public Utilities Commission [CPUC]. 2006. *California Energy Efficiency Evaluation Protocols: Technical, Methodological and Reporting Requirements for Evaluation Professionals*. San Francisco, CA: California Public Utilities Commission.
- Efficiency Valuation Organization [EVO]. 2010. “Certified Measurement and Verification Professional Certification” and “Fundamentals of Measurement & Verification: applying the New IPMVP” training courses; website: www.evo-world.org/index.php?option=com_content&task=view&id=278&Itemid=337.
- Hall, N. R. Ridge, G. Peach, S. Khawaja, J. Mapp, B. Smith, R. Morgan, and P. Horowitz. 2008. “Reaching Our Energy Efficiency Potential and Our Greenhouse Gas Objectives – Are Changes to our Policies and Cost Effectiveness Tests Needed?” *Proceedings of the 2008 AESP National Conference*, Association of Energy Services Professionals, Phoenix, AZ.
- Hewitt, D. 2000. “The Elements of Sustainability,” *2000 ACEEE Summer Study on Energy Efficiency in Buildings*, pp. 6-179 – 6-190, American Council for an Energy-Efficient Economy, Washington, DC.
- Hirst, E. and J. Reed (Eds.). 1991. *Evaluation Handbook*, ORNL/CON-336, Oak Ridge National Laboratory, Oak Ridge, TN.
- National Action Plan on Energy Efficiency [NAPEE]. 2007. *Model Energy Efficiency Program Impact Evaluation Guide*, available at: www.epa.gov/cleanrgy/documents/evaluation_guide.pdf
- Rambo, E. and B. Ward. 2009. “Focus on Energy Evaluation – Emission Factors Update,” PA Consulting Group, Madison, WI.
- Skumatz, L., M. Khawaya, and J. Colby. 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,” available at: <http://uc-ciee.org/energyeff/energyeff.html>.

TecMarket Works. 2004. *The California Evaluation Framework*, available at: http://www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf.

Titus, E. and J. Michals. 2008. "Debating Net Versus Gross Impacts in the Northeast: Policy and Program Perspectives," pp. 5-312 - 5-323, *2008 ACEEE Summer Study on Energy Efficiency in Buildings*, pp. 5-219 – 5-234, American Council for an Energy-Efficient Economy, Washington, DC.

U.S. Department of Energy [USDOE]. 2007. *Impact Evaluation Framework for Technology Deployment Programs*, available at: http://www.eere.energy.gov/ba/pba/km_portal/docs/pdf/2007/impact_framework_tech_deploy_2007_main.pdf.

Vine, E., R. Pahl, S. Meyers, and I. Turiel. 2009. "An Approach for Evaluating the Market Effects of Energy Efficiency Programs," *Energy Efficiency*, 2009.