

The Tangled Web of Evaluation in the Northeast United States: A Program Administrator Perspective

Jeremy Newberger, National Grid, Waltham, MA

Abstract

Within the past five years, many changes have begun to take place in the oversight and management of energy efficiency evaluation studies in the northeast United States. This is motivated by the increase in funding for energy efficiency programs from traditional sources such as system benefit charges or ratepayer funds, as well as new sources of funds such as Federal stimulus grants, emerging capacity and carbon markets in which energy efficiency participates or plays a role, and others. With this increase in funding have come increased scrutiny, direction, and oversight on how evaluation is being done.

For program administrators, the new structures provide a confusing array of requirements. Evaluation practices that are acceptable in one state are not acceptable in others, or in regional capacity markets. Studies that could be economically performed as multi-state efforts are less likely to be supported due to these varying requirements. Even though program administrators still have responsibility and accountability in regulatory proceedings, they are ceding authority to other parties that have oversight mandates. The result of all this is a tangled web of evaluation frameworks.

The background to evaluation frameworks in the northeast United States is provided. Measures of success in the conduct of energy efficiency evaluations are defined. In this context, the emerging web will be described and reviewed. The examination will conclude with some recommendations for improvement in the regional evaluation framework that have the potential to satisfy many of the competing interests. This analysis should also be of interest to other states, regions, and countries that feature multiple energy efficiency programs, policies, funding sources, and evaluation stakeholders.

Introduction

The northeastern portion of the United States, see Figure 1, comprises states from Maine to Maryland, and includes Washington, D.C.¹ It is generally considered to be one of the progressive areas of the country – all of the jurisdictions voted for Barack Obama in the 2008 presidential election. Many of the states in the region are in the forefront of energy efficiency efforts in the United States. Some states have had programs for more than 20 years. In recent years, many of the states in the region have passed legislation specifying aggressive targets for efficiency.

The main entities involved in the delivery of energy efficiency have been electric utilities acting as program administrators. In some of the less populated states with smaller utilities, state agencies themselves have assumed a lead role in the delivery of energy efficiency. Utility program delivery is regulated by utility commissions. Like other areas of the country, energy efficiency programs must achieve standards of cost effectiveness and equitability in program offerings.

¹ New England is comprised of Maine, New Hampshire, Vermont, Massachusetts, Connecticut, and Rhode Island. The Mid-Atlantic states are Pennsylvania, Maryland, New Jersey, and Delaware. New York is not considered to be part of either region.

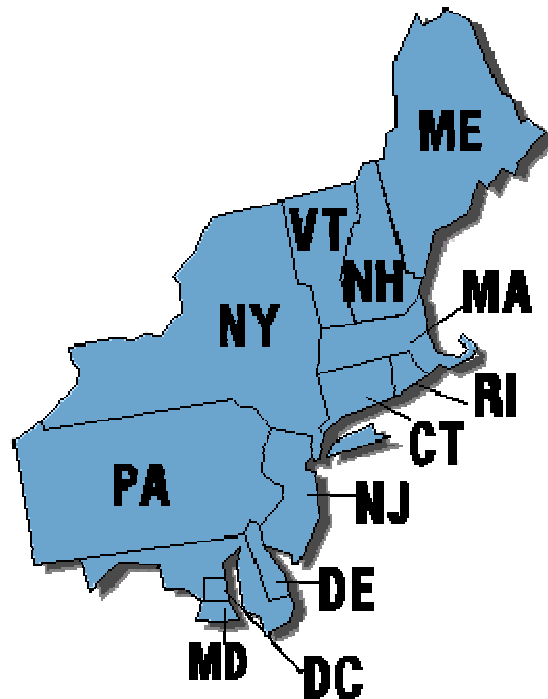


Figure 1. Northeast United States

Ten years ago, energy efficiency program administrators in New England had full responsibility for program evaluation. Process and impact evaluations were most commonly performed by contractors hired by the program administrators. The work of the contractors was supervised, and reports were approved, by program administrator staff. A loose array of non-utility party consultants occasionally were involved in particular studies of interest. Public utility commissions (PUCs) were provided with the evaluation findings and studies and could investigate either but only rarely did.

Within the past five years, many changes have taken place in the energy efficiency marketplace. There are emerging capacity markets² in which energy efficiency is allowed to participate. A regional carbon market has been started. Both of these markets, along with Federal stimulus funds, have increased the funding for energy efficiency. On top of that, many states have created ambitious energy efficiency procurement mandates, and promised even more funding. In this dynamic and growing environment for energy efficiency, more attention is being paid by more parties to measuring savings and the detailed oversight of evaluation.

States are providing more direction and imposing more requirements on evaluation. Capacity market operators are specifying evaluation protocols to ensure that energy efficiency capacity is reliably measured in a manner consistent with the way generation output is measured. Regional evaluation collaborations have been constituted in response to these capacity market requirements. New state oversight boards have, in the interest of increasing transparency and accountability, imposed new oversight structures on program evaluation which leaves day-to-day evaluation study management in the hands of the program administrators, but removes the role of final decision making from them.

² ISO-New England and the PJM Interconnection, power grid operators in New England and the Mid-Atlantic States, respectively, have created capacity markets where they identify the capacity requirement three years in the future. Qualified supply and demand side projects bid to supply that capacity. Market rules have been in place since 2007 and ISO-NE's first auction was in February 2008. Winning bidders from the first auction are to deliver their capacity to the grid June 1, 2010. ISO-NE's fourth auction will be in August 2010 for delivery of capacity in June 2013.

For program administrators, the new structures provide a confusing mix of requirements. Evaluation practices that are acceptable in one state are not acceptable in others. Studies that could be economically performed as multi-state efforts are less likely to be supported. Even though program administrators still have responsibility and accountability in regulatory proceedings for evaluation results on which claimed savings are based, they are ceding decision-making authority to other parties that often do not have regulatory authority. The result of all this is a tangled web of evaluation frameworks. It is unclear whether the outcome of these developments will provide consistent and statistically defensible results, or will be transparent, independent, or economical. What is clear today is that the different parties involved in developing evaluation rules are not considering the impact of their decisions on other users of evaluation results.

Implementation and Evaluation Frameworks in New England

Implementation of energy efficiency programs in the Northeast is currently done under a few different models. There are several states with utility program administrators (PAs), such as Massachusetts, Rhode Island, Connecticut, and New Hampshire. Vermont and Maine have state-sponsored efficiency program administrators (Efficiency Vermont and Efficiency Maine Trust). New York is a hybrid model where a state agency, the New York State Energy Research and Development Authority (NYSERDA), and utility program administrators run programs that are intended to be complementary.

Under each one of these frameworks, the evaluation of energy efficiency savings and impacts is performed. Table 1 summarizes the different dimensions of evaluations in the seven states. In this table, the fields are defined as follows:

- Who Manages Evaluation? – In most cases, evaluations are performed by outside firms. This field identifies, on a day-to-day basis, who manages the evaluation studies.
- Who Oversees Evaluation? – Who has decision-making authority over what studies are done, reviews methodologies and interim results, and determines when studies are completed, accepting the results as final?
- Who Approves Evaluation? – Who passes judgment on the application of evaluation study results to efficiency programs?
- Standards of Review – When the approver passes judgment, what standard of review do they apply? A reasonableness review is a general review that the application of results makes sense. High Level and Detailed Technical Review means that the study results are reviewed at some level of detail.

Table 1. Summary of Current State Evaluation Frameworks in the Northeast United States

State	Who Manages Evaluation?	Who Oversees Evaluation?	Who Approves Evaluation?	Standards of Review
Connecticut	Program administrators	Energy Conservation and Management Board (ECMB) (oversight board)	Energy Conservation and Management Board, with PA input	High level review of studies by consultant to ECMB
Maine	Program administrator	Maine PUC	Maine PUC	Varying levels of technical review by Maine PUC
Massachusetts	Program administrators	Energy Efficiency Advisory Council (EEAC) (oversight board)	MA Department of Public Utilities (regulator)	Varying level of technical review of studies by consultant to EEAC
New Hampshire	PAs conduct most studies	NH PUC	NH PUC	Reasonableness review
New York	Program administrators	NY Public Service Commission	NY Public Service Commission	Too soon to tell
Rhode Island	Program administrators	Program administrators recommend plan to RI PUC	Program administrators and Division of Public Utilities and Carriers	Reasonableness review by Division of Public Utilities and Carriers
Vermont	VT Department of Public Service or its contractors	VT Department of Public Service	VT Department of Public Service and Efficiency Vermont (PA); VT PUC arbitrates, if necessary	Reasonableness review to detailed review of studies, depending on study

Measures of Success in Evaluation

Under these different structures, many measures or indicators of success are used to guide evaluation efforts. The primary indicator of a successful study is obviously one that satisfies the objectives of the particular study. The objective of a study may be as simple as “Measure and verify the energy savings of the ENERGY STAR Homes program implemented in 2009.” Secondary indicators of success are those that describe the environment in which evaluations are conducted; they shape and guide evaluation design, management, and execution. Unlike the primary objective, these measures are applicable to a wide range of studies.

- Statistical measures of success are confidence and precision. These are used by both evaluators and regulators/reviewers of evaluation studies. They provide an indication of the robustness of the results. The confidence and precision goals are often set by oversight bodies, who are very interested in study design, but this has no influence on actual study results. Most studies usually focus on kWh savings and are designed to have a 90% confidence with a +/- 10% relative precision. However, the Independent System Operator – New England (ISO-NE) focuses on kW and requires 80% confidence /10 % relative precision on delivered kW. The delivered kW is usually a function of the bundling of savings across a group of measures and programs and,

consequently, is dependent on the results of a number of evaluation studies associated with the programs.

- Societal success of evaluation is whether the evaluation study demonstrated achievement of a certain amount of energy savings or a certain amount of CO2 emissions reduced. By this standard, a successful evaluation study would be one that could be used to conclusively determine if the program met or did not meet its goals, such as energy savings or emission reductions. The success of the study in this regard is therefore independent of whether the results are “good” or “bad” for the program.
- Economic measures of success indicate whether evaluation resources have been used efficiently. These measures may be the overall cost of the study, cost per site, cost per survey, or cost relative to the amount of savings that might be affected by the study (an indicator of cost effectiveness). The dedicated hours spent by the evaluator, support staff, or customers are also indicators. These measures are used most frequently by evaluators in bid review and in study management, but are not given much consideration by reviewers or regulators. Overall budgets for evaluation are considered (usually as a percentage of the overall budget for energy efficiency programs) by oversight agencies. However, where in the past evaluation efforts might have been combined across states to create economies of efforts, there is a growing balkanization of evaluation in the interest of obtaining state-specific results. Consequently, as long as evaluation budgets are close to the target amount of evaluation spending, there is little interest in economic efficiency in evaluation planning, which may result in cost inefficiencies.
- Regulatory acceptance is a clear measure of success. In this context, regulators are the entities that authorize cost recovery, program administrator performance incentives, or other financial payments based on documented program savings. They include both PUCs as well as market operators who operate capacity or carbon markets in which energy efficiency is eligible to participate. However, regulatory staffs often do not have sufficient technical resources to perform detailed reviews of evaluation studies. Regulatory agencies instead indicate acceptance by conducting a thorough reasonableness review of the application of results from evaluation studies. In this function, there is an expectation that the documentation of savings is provided and clearly linked to studies. In addition, in the case of ISO-NE, a rule-based system has been developed where market participants must certify adherence to a number of rules (precision, confidence, measurement accuracy, age of studies) related to demonstrating energy efficiency savings.
- Political success is obtained most frequently through the acceptance by oversight bodies. These bodies are concerned with the appearance of objectivity, transparency, and independence in the conduct of evaluation studies. As stated in the resolution of the Massachusetts Energy Efficiency Advisory Council establishing new guidelines for measurement and evaluation oversight in that state, “There is a need to ensure both the reality and the perception of the independence and objectivity of EM&V activities, as well as the need to help ensure consistency, timeliness, and credibility of the results” (MA Plan 2009). In the political context, the appearance of objectivity is just as important as actual objectivity.
- Professional success is an ephemeral but important measure. The evaluation community is fairly small (compared to implementation), and some practitioners in evaluation firms or program administrator staff have years of expertise and experience. They are motivated by getting to the truth behind program performance and feeding the findings into program reporting as well as continuous program improvement. Evaluation practitioner success is measured by their contribution to these elements, and the regard that others have for their expertise and judgment.

The Emerging Web of Evaluation Frameworks

Within the past two years, there has been an increased state-centered focus in evaluation activities. This focus is motivated by increasing expenditures on energy efficiency and new state rules regarding program oversight and implementation. With the increased expenditures, oversight bodies and regulators are imposing more evaluation requirements. Accountability concerns dictate that there be an alignment between the program administrators and the oversight or regulatory bodies in that state. There is also a strong interest to implement changes quickly in the evaluation frameworks concurrent with the ambitious energy efficiency plans. At the same time, yet another confounding factor within the region is the influence of the recession on state economies, so that at the same time that programs are ramping up and demands on measurement and verification (M&V) are increasing in some states, energy efficiency budgets are threatened and infrastructure is being dismantled in other states (NEEP Policy). There is a lot of upheaval in a segment of energy efficiency that is used to being fairly stable.

The “state focus” results in a decreased ability to pool evaluations across states for those program administrators who operate in different states (there are several of these in New England and New York), or traditional cross-program administrator cooperation. This pooling had been done because programs in different states were considered to be homogeneous and resulted in cost efficiency in evaluation. This consideration has been overridden by the interest in more state-focused evaluation.

The oversight bodies pose a particular challenge. Oversight boards not affiliated with regulatory agencies exist in Massachusetts, Rhode Island, and Connecticut. These boards have oversight over energy efficiency planning and, in Massachusetts and Connecticut, over evaluation as well. (In Rhode Island, the board has an advisory role). These bodies’ opinions feed into the regulatory process, but they are not accountable for the results. Program administrators are still accountable before regulators for savings claims based on evaluation results. Ideally, program administrator staff working with the oversight bodies will resolve all differences regarding the results of evaluation studies prior to their use in regulatory proceedings. However, there may be cases where program administrators will be obligated to justify study results that they felt were not acceptable but were compelled to use by the oversight board.³

At the same time that there has been an emerging state focus on evaluation, regional forces have also been emerging, chiefly in the area of capacity markets where demand resources, such as energy efficiency, are allowed to participate. Both the New England market, run by ISO-NE, and the Pennsylvania-New Jersey-Maryland (PJM) Interconnection (serving the mid-Atlantic and parts of the Midwest), have allowed energy efficiency to compete alongside supply resources. These are extremely complicated markets, but the general rules for the participation of energy efficiency – including those of evaluation and measurement – center around treating demand resources as though they were supply resources to the extent possible, and focus on the proving the reliability of the capacity provide through demand savings. ISO-NE created a manual for the “Measurement and Verification of Demand Reduction Value from Demand Resources” to meet its needs (ISO-NE).

The manual’s recommendations in a few areas for the treatment of demand side capacity are noteworthy. For instance, the output of supply resources is measured by revenue grade meters with very narrow tolerances. For the measurement of energy efficiency savings, ISO-NE has specified accuracy of

³ Evaluation professionals may have differing opinions on methodology, interpretation of results, or the proper use of preliminary results that are expected, but are nonetheless preliminary. Even where there are strong arguments on both sides, the oversight board will now have the final say. In addition, program administrators also may be under pressure to resolve an issue and accept results they would otherwise not accept due to time pressures associated with filing deadlines and the threat of intervention that may delay filing preparations. Many of these factors were involved in a case in Massachusetts in 2009 (prior to the adoption of the new evaluation framework) regarding the use of preliminary results from a recent study that became available in 2009 at the time program administrators were preparing their reports on energy savings for 2008.

metering equipment that was beyond the capability of a significant percentage of the metering equipment used by energy efficiency program evaluators. Evaluators had to replace a portion of their meter stock in order to comply. Also, since the energy efficiency savings are a result of the measurement of many parameters (for example, voltage, current, percent load), measurement rules require that the chained product of these parameters have an overall accuracy of +/- 2%, which means that the accuracy of the components must be even more accurate. ISO-NE also requires that evaluation results used in the verification process must be less than five years old, or their use must be justified. Finally, for now, ISO-NE is satisfied with demand savings results that meet the 80/10 standard at an aggregated portfolio level. Should they require 80% confidence and 10% relative precision of results for a geographic area, for a program, or for a study, it will increase sample sizes and evaluation costs. The PJM Interconnection has also developed M&V requirements in support of their capacity market. These requirements are similar in many instances to ISO-NE's.⁴

Due to some of the regional forces shaping M&V activities in the capacity market, a regional Evaluation, Monitoring and Verification (EM&V) Forum was constituted in 2008, under the auspices of the Northeast Energy Efficiency Partnerships (NEEP Forum). The Forum is made up of 11 member states in the Northeast plus the District of Columbia. The Forum conducts studies of regional interest – such as those focused on methodologies or on determination of coincidence factors – that support participation in the capacity markets. The Forum operates under a Steering Committee made up of regulators from the member states and is staffed by NEEP. Program administrator personnel work on topic committees and management of specific studies. However, while associations of utility Commissioners have affirmed the goals of the Forum (NEEP site), and the Forum Steering Committee has held discussions about what it means for the Steering Committee to adopt a Forum product, no explicit connection currently exists between a state regulator's approval of a particular study through the Forum and the program administrator being required to use the results of a study; participation is voluntary on a study-by-study basis. Nor is there reconciliation between the regional focus of the Forum and the growing state-specific focus or the state regulatory process involving energy efficiency programs.

On a national level, the North American Energy Standards Board (NAESB) is working on developing a national EM&V standard (NAESB). ISO-NE's staff who are involved in the regional capacity market M&V requirements are also involved in this effort. There is a high likelihood that national standards will be similar to ISO-NE's M&V requirements. On the other hand, none of NAESB's retail electric or gas members are Northeast U.S. electric or gas energy efficiency program administrators, and only Massachusetts regulatory agencies are listed on the NAESB website as being regularly involved in NAESB activity. Several program administrators in the PJM region participate in NAESB. This raises the question of whether this national standards organization has the standing to create standards that would apply to efficiency programs in the Northeast. Alternatively, if NAESB creates standards, will state regulatory commissions that are not involved in NAESB adopt them for their own state standards? It is possible that adoption of the yet-to-be-created NAESB standards would be done on a state-by-state basis, further increasing the complexity of the evaluation landscape.

Another national effort is being undertaken by the State Energy Efficiency (SEE) Action Network (formerly known as the National Action Plan for Energy Efficiency, or NAPEE). This effort, launched under the auspices of the U.S. Department of Energy and the U.S. Environmental Protection Agency, is designed to enhance energy efficiency efforts in the 50 states by creating a national support system to foster the development of energy efficiency. One of the focuses of the effort has been on the creation of resources regarding measurement and verification (NAPEE). This effort is not yet at the stage of developing protocols; it is currently focused on cataloguing measurement and verification

⁴ PJM did not adopt the +/- 2% accuracy requirement for chained results.

activities and identifying needs to support state activities. Thus far, the initiative has identified three main areas of need: improving the methods used to estimate energy efficiency program energy savings; improve quality control and accuracy of reported program savings estimates; and improve the consistency in defining program energy savings terms and in reporting energy savings.

Rating the Emerging Web

As the web of evaluation frameworks emerges, it is useful to review the “measures of success” outlined above to observe how the web fares and perhaps may be improved.

Statistical

It is too soon to tell what impact the emerging web will have on the statistical measures of success. Thus far, interested parties, other than ISO-NE, have not tinkered with the common practice of 90/10. What is not clear is whether and how the gap between ISO-NE’s 80/10 standard on results will be reconciled with the 90/10 standard in design used in other forums, or whether the ISO-NE’s focus on kW will interfere with the broader focus on kWh savings. It is possible that statistical measures of performance will continued to be achieved. One consideration will be whether, with all of the evaluation studies that are to be performed across the region, it will be too expensive to maintain a 90/10 standard in evaluation study design, or approach that standard in results. A looser standard may be deemed acceptable in order to manage costs and other resources.

Societal

It is also too soon to tell whether the evaluation studies will be successful in demonstrating the actual success towards achieving societal energy savings or carbon reduction goals. A critical mass of completed studies is needed to begin to ascertain success in this area. The fundamental elements of evaluation studies (e.g., sampling, measurement techniques) are largely unchanged, so it is reasonable to believe that valid results will be obtained. What may complicate the determination is whether some of the shortcomings noted here – such as resource constraints or differences of opinion about interpretation of results – will interfere with the determination of societal success.

Economic

Early indications are that the web of evaluation is more costly than the previous program administrator-led effort. Using Massachusetts as an example, statewide evaluation budgets totaling \$60 million over three years have been approved. National Grid’s evaluation budget in Massachusetts has increased from \$4 million in 2009, to almost \$6 million in 2010 and \$11.4 million in 2012. The increase in evaluation funds is steep. However, as a percentage of its program budgets in Massachusetts, National Grid’s evaluation costs are fairly stable – between 3.5% and 3.8%. This is because in budget planning, evaluation was assigned a target percentage of the budget.

The increase in costs is due in part to an increase in the amount of energy efficiency being implemented and to process evaluations focusing on new programs and initiatives. However, the new structure currently lacks any focus on achieving economies in evaluation efforts. In Massachusetts, part of the cost is due to a lack of specificity in the blanket RFPs that have been issued to date. Competitive bidding does not seem to have affected study pricing. Bidders are providing conservative cost estimates

for relatively vague scopes of work. There will be an opportunity to hone these costs as work is performed.

There is also a likelihood that the increased state focus will lead to a duplication of effort of doing the same study in several states, or performing joint studies with increased sample sizes so that statistically valid results for each state may be obtained. It is premature to assess the cost of this change as smaller states that had previously joined studies that were designed on a regional basis begin to adapt to having to do their own studies or add full sample complement for their state to another states' study. One goal of the NEEP Forum is to be able to reduce some evaluation costs within the region by leveraging resources. However, that economic opportunity is somewhat challenged by virtue of the tension between state-specific changes and the emergence of regional framework.

Regulatory

Thus far, the only studies emerging from the web that have been completed are studies done under the NEEP EM&V Forum (and program administrator studies launched under the old regime). While the authority of its Steering Committee to influence activities in specific states is somewhat clouded, NEEP has taken pains to include both utility and environmental regulators in its steering committee. The structure and collaborative environment within the Steering Committee and the feedback loop between Forum and regulators allows regulators to be informed and engaged. It remains to be seen whether regulatory acceptance at the Forum level translates into action at a state level.

Outside of the Forum, regulators are likely to be satisfied with the added state-specific oversight and appearance of independence provided by the new framework. However, where there is an interest in minimizing administrative costs, regulatory bodies may be unhappy about the significantly higher cost of evaluation compared to previous years.

Political

Oversight bodies and other political entities are generally pleased with the evaluation web, thus far, because it has provided them with more control – or the appearance of control – over evaluation. They have endorsed broad evaluation plans which incorporate the features they desire. (MA Plan) This dynamic bears watching as studies progress and the realities of study management involving multiple program administrators and overseers occur: delays, data difficulties, differences of opinion, and limitation of resources are encountered. The more studies there are, the more the oversight bodies will be challenged to perform their oversight and offer opinions on issues that will arise on each study. It is likely that the oversight bodies will not have sufficient resources to oversee management of the studies for which they have control. It is possible that study timetables, study priorities, or expectations about the relationships between study managers and study overseers will need to be re-examined by the parties (oversight bodies and program administrator evaluation staff) that agreed to these elements.

Professional

Already, evaluation professionals are feeling the strain of the new framework on two fronts. The new framework requires additional human resources. It is not just that there is more oversight, but also there is more work; this is a problem for program administrators, regulators, oversight boards, and staff at consulting firms engaged to perform studies. Program administrator staff are adjusting to – and accepting – having more overseers, higher workloads and budgets, and of established practices being questioned. However, these practitioners are also being told that, because of the need for independence, their professional experience and judgment are not as highly regarded as they once were.

Second, evaluation consulting firms are finding it necessary to join forces with other firms to be able to marshal sufficient resources to handle large scopes of work. Nevertheless, some evaluation firms are finding their resources are somewhat constrained, and some firms are finding it necessary to decline to bid on some studies largely because of resource or time constraints. Everyone is stretched too thinly to take a step back to advocate for changes that would serve common evaluation priorities. This poses an opportunity that could be developed but, for now, it is not a priority. Instead, the evaluation community finds itself in reaction mode in an attempt to reconcile the demands and results of similar studies in different jurisdictions.

In summary, the emerging web is shaping up to be politically positive, economically negative, statistically neutral, and professionally challenging.

Areas for Improvement

Thus far, this paper has described the emerging framework for efficiency program evaluation in the Northeast U.S., and the difficulties that the framework poses. However, there are some potential areas for improvement that could be implemented in the region, or by others who might face themselves in similar situations.

First, all involved should accept that, while there is general agreement about what energy efficiency is, it still means different things to different audiences. Energy efficiency is: a capacity resource; an economic stimulus vehicle; an income redistribution tool; our best weapon against climate change; a business decision; and a source of energy savings. The perspectives of different interest groups add to the entanglement. If the different perspectives on energy efficiency are sorted out, a consistent evaluation framework could be developed. Climate change regulators have noted that a kWh should be the same in all jurisdictions, but it is not. There are variations in baselines and program attribution that are important to energy efficiency program regulators, but are of less importance to parties that are only interested in changes in consumption. One solution may be for two sets of energy savings to be tabulated under uniformly accepted rules: e.g., gross energy savings and net energy savings. Parties that are interested only in gross savings (e.g., capacity market operators) would receive consistently defined and calculated gross savings. And parties that were interested in net savings (e.g., regulators in some states) would be provided with net savings.

Second, ISO-NE and other grid operators have created stringent measurement standards for efficiency so that it could be treated like supply. Those rules should be modified to accommodate the different qualities of energy efficiency, such as the reliability created by the diversity of hundreds of installations as opposed to the reliability associated with the precision of measurement. If an efficiency supplier cannot meet the 80/10 standard, perhaps the demand savings for which it is credited should be discounted accordingly, as opposed to being rejected completely. Efficiency suppliers should be given the option of trading the precision of specific measurements for the security of large numbers of efficiency installations in for ensuring reliability.

Third, from an economic perspective, it is imperative to prove or disprove the concept that state programs are sufficiently heterogeneous as to warrant state-specific evaluations. Millions of dollars could be saved in the long term by eliminating oversampling or duplicate studies if state-specific studies are not blindly required.

Performing one large evaluation across several states should be less expensive than doing a similar study for each state. It would be a good idea to test this hypothesis. This is particularly appropriate for impact evaluations which measure the realization of energy savings compared to predicted engineering estimates. This recommendation may be less applicable to market effects or

process evaluations where, despite similar program designs, there are differences in the way the program is implemented, depending on who is administering or delivering the program. Differences in baselines across states may have to be studied to see if they are significant in the design of studies. And, to the extent that different regulatory policies are driving evaluation design, such as requirements on gross versus net savings, it may be an opportunity to take a fresh look at whether such policies should be homogenized.

Fourth, with so many overlapping jurisdictions, there is a strong need for coordinated decision making. In the Northeast U.S., the NEEP EM&V forum may be best suited for this role because it already spans the region. In addition, its focus on methodological or other studies of regional interest is highlighting the similarities across states, rather than creating divisions.

The potential for the NEEP Forum to assume this role should be considered, with attention to practical issues that would need to be worked out. Foremost among these issues are NEEP's interest and ability to assume this role. The EM&V Forum's regulatory relationships are still developing. Regulators on its steering committee are not binding program administrators in their respective states to its decisions. Instead, each state or organization is making its own decision regarding study acceptance. NEEP, its sponsors, and the Forum Steering Committee should explore avenues – whether through assuming a more overt evaluation policy coordination role or perhaps in regulatory proceedings – for it to become the platform for clear regional direction on M&V.

Similarly, regional entities could be developed and come together on a national basis to deal with common issues. SEE might be a good forum for a national effort as it is more organic and inclusive than the NAESB effort. While it might be argued that national standards should be mandatory, voluntary and participatory efforts that focus on developing consistent standards which are less costly to implement may be more widely accepted and consequently more effective.

Conclusion

As described, the current tangled web of evaluation is not a creation of a single designer. It is a result of many changes that have shaped the landscape for energy efficiency and, indeed, are even helping program administrators reach unprecedented levels of energy savings. By several measures of success, continuing to conduct evaluations under this framework will provide results that are valid, useful, and defensible. By other measures, chiefly relating to the efficient use of evaluation resources, the framework is suboptimal.

Given the amount of coordination needed and the challenging issues to resolve, the improvements recommended here to the regional – and even national – evaluation framework may take years to implement. However, such an effort will be well worthwhile. It could lead to a framework that would be more flexible, reasonable, and consistent while still satisfying the drivers that have increased the focus on M&V and the demand for transparency and accountability. These improvements would untangle the web of evaluation. At the same time, an improved framework would support the continued advancement of energy efficiency as a vital factor in the economy, and a reliable tool in the campaign to reduce dependence on fossil fuels and slow the forces of climate change. In such an environment, it would not matter what entity is driving the evaluation or how the evaluation is being done or what geographic area is being studied; the results would always be comparable.

Acknowledgments

The author acknowledges the thoughtful review provided by David Jacobson of National Grid, and Elizabeth Titus of the Northeast Energy Efficiency Partnerships. The author takes full responsibility

for the contents of this paper. All opinions expressed in this paper do not necessarily state or reflect those of National Grid.

References

- [ISO-NE] ISO-New England. *ISO New England Manual for Measurement and Verification of Demand Reduction Value from Demand Resources: Manual M-MVDR*. http://www.iso-ne.com/rules_proceeds/isone_mnls/MMVDR/m_mvdr_measurement_and_verification_demand_reduction_revision_0_04_13_07.doc
- [MA Plan 2009]. *2010 – 2012 Massachusetts Joint Statewide Three-Year Electric Energy Efficiency Plan..* <http://www.ma-eeac.org/docs/DPU-filing/ElectricPlanFinalOct09.pdf>, page 275
- [NAESB] North American Energy Standards Board.
Standards development at http://www.naesb.org/retail_request.asp;
Membership at http://www.naesb.org/pdf4/members_quadrant.pdf
- [NAPEE] National Action Plan on Energy Efficiency. 2007. *Model Energy Efficiency Program Impact Evaluation Guide*, available at: www.epa.gov/cleanrgy/documents/evaluation_guide.pdf; also see www.epa.gov/cleanenergy/energy-programs/napee/index.html
- [NEEP Forum] Northeast Energy Efficiency Partnerships. <http://neep.org/emv-forum/about-emv-forum>
- [NEEP Policy] Northeast Energy Efficiency Partnerships. <http://neep.org/public-policy/poc-states/poc-states-ct> and <http://neep.org/public-policy/policy-outreach-and-analysis/poc-states/poc-states-nj>