The case for expanding comprehensive US evaluation mandates and considerations for the EU

Gary Epstein
ERS
120 Water Street
North Andover, MA 01845
USA
gepstein@ers-inc.com

Jeff Perkins
ERS
jperkins@ers-inc.com

Jon Maxwell
ERS
jmaxwell@ers-inc.com

Lucy Neiman
ERS
lneiman@ers-inc.com

Patrick Hewlett
ERS
phewlett@ers-inc.com

Sue Haselhorst
ERS
shaselhorst@ers-inc.com

Keywords

CO_2 reduction, evaluation, evaluation methods, monitoring, impact evaluation, process evaluation

Abstract

With new mandates for efficiency largely associated with the push for carbon emissions reduction, but also driven by the need for energy and economic stability, installed efficiency projects must be able to ensure energy savings and be fully defensible under increasing scrutiny. In the US, comprehensive impact and process evaluations have been a mainstay in the energy efficiency industry for many years, but the quality and rigor of evaluations can vary considerably from state to state and from program to program. In the EU, while mandates for efficiency are strong, a concurrent call for third-party evaluation has not been as evident; the need for evaluation is acknowledged, but the responsibility is often left in the hands of the project implementers. While that approach may often be successful, it at least should be stated that, like the US, evaluation approaches may be inconsistent between EU member states.

This paper will discuss the rationale for technically sound and consistent evaluation protocols, focusing on the US systems and recent methodology evolution and enhancements. Through incorporation of appropriate, cost-effective evaluation protocols, efficiency projects will be viewed as progressively dependable, and the efficiency industry will successfully remain a mainstay in the quest to resolve climatic and other energy problems. The paper will provide a comprehensive discussion of the most common US-focused standard evaluation approaches and protocols as well as enhanced evaluation approaches that can help to ensure cost-effectiveness for different and evolving program structures and budgets. Finally, we will briefly discuss the potential of applying US-evaluation approaches to the EU market with the objective of motivating conversation regarding the potential need for consistency in approaches for evaluations performed in the US and EU.

Introduction

For both the US and EU, and for all the US states and EU member states, there must be a clear mandate for evaluation efforts with an approach that demonstrates quality and consistency in the assessment of program performance and the determination of energy impacts. Further, where appropriate, there is considerable virtue in, as well as a need for, independent third-party evaluation. Any entity that has an energy project performance goal is potentially biased. This includes, roughly in order of most to least bias potential: ESCOs with performance guarantees, equipment vendors, program contractors, program administrators, customer project advocates/implementers, customer decision-makers, and policymakers. The only way to eliminate the possibility of bias is for an independent third party to assess the results of the work.

While there are broad definitions of evaluation, for the energy efficiency project and program world, these typically fall into two categories – impact evaluations and process evaluations. While they are often performed as distinct activities or efforts, their common objective is to assess the overall performance of programs and projects, basically asking the questions, respectively, "Have the projects or programs had the impacts that was intended and estimated?" and "Have the programs and projects been handled effectively from a procedural perspective, representing energy users and associated market actors?"

IMPACT EVALUATIONS

Impact evaluations tend to handle the more quantitative side of the field, with an end objective of determining the energy (kWh) and demand (kW) impacts achieved on both project-specific and whole-program levels. Conducting an impact evaluation is most typically performed as an engineering effort involving site visits to inspect project efficiency equipment or controls, verifying that the systems have been installed as anticipated and are operating as intended. Data is gathered by evaluation engineers and, depending on the scope of that evaluation, metering equipment might be deployed to record actual operational characteristics. All site-gathered data, along with metering data as applicable, is used in engineering calculations to determine the actual usage pattern of the efficient system. This is then compared to the baseline system energy use to develop values for the project savings or impacts for that site. These evaluated savings might differ from pre-installation estimates, and the ratio of the evaluated estimate to the pre-installation estimate is referred to as a site or project realization rate.

To ensure the cost-effectiveness of an evaluation effort, particularly for efficiency programs that have many project sites (often hundreds or thousands), only a sample of participants (rather than a census) is subjected to the scrutiny of evaluation site visits. Ultimately, the evaluation effort will take the results of each sample participant and weight those up to produce program level impacts (energy and demand) and realization rates. These are a measure of how effectively pre-estimates of potential project energy savings have been and should be helpful in improving the results of future projects.

There are many other aspects to impact evaluation projects. One area of consideration that is often addressed for US program evaluation is associated with attribution, i.e. describing and quantifying causality of energy savings. For US programs, it is often necessary to determine whether it was the efficiency program and its financial incentives that truly motivated the project and its energy savings or whether the end-use customer would have acted in the same way without program intervention. The result of an attribution assessment is to adjust (generally downward) the previously developed gross energy impacts and realization rate, thereby producing net results.

PROCESS EVALUATIONS

Process evaluations are focused on characterizing the overall effectiveness of an energy efficiency program, looking at topics such as organization of program staff and the logic of their procedures, customer satisfaction, engagement of various market actors, and success of the program in reaching various end user groups. Comprehensive process evaluations will include benchmarking efforts, assessing the overall success of the program and benchmarking it against programs that are considered best-practice models.

It should be noted that the results of process and impact evaluation results are closely related. Programs that are found to have procedural deficiencies or limitations will often have lower realization rates. While impact evaluations are often treated as distinct and independent efforts, many of the tasks for these projects are so interrelated that there are great reasons to perform integrated impact and process evaluations. This is discussed in detail in a subsequent section of this paper.

PREVALENCE OF EVALUATION IN THE US

Efficiency programs in the US have represented a considerable effort in many progressive states for 2 to 3 decades. Such programs are typically mandated at the state level and are funded through fees paid through energy bills. These programs provide a wide variety of services, including marketing and direct outreach; technical measure development support; engineering design assistance; and project implementation incentives. Along with these program mandates is the mandate for comprehensive evaluation (impact and process) of these programs. As such, energy project and program evaluation in the US is a mature focus of the energy industry and involves many highly experienced third-party practitioners. While evaluation approaches and scope are constantly evolving, it is rare that major program efforts do not receive the scrutiny and quality assessment of mandated evaluations.

EVALUATION IN THE EU

In contrast, while EU state efficiency mandates are growing, comprehensive scrutiny and standardized evaluations have not been as common as in the US. The *Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on Energy Efficiency* is the primary overarching legal document that recognizes the need for energy efficiency actions to curb emissions and makes binding commitments of EU members to do so. The Directive's Annex V Common methods and principles for calculating the impact of energy efficiency obligations specifies that member states will notify the Commission of detailed methodology for, among other things, "monitoring and verification [evaluation] protocols and how the independence of these from the obligated, participating or entrusted parties is ensured." Clearly, there is now a mandate for evaluation of energy efficiency projects that are being implemented as part of the EU effort to address climate change. Further, standardized, acceptable, and effective protocols must be adhered to in conducting these evaluations. Further, it is noted that the National Energy Efficiency Action Plans mandated for each state may specify evaluation approaches to be utilized in their jurisdictions. While comprehensive impact and process evaluations may not be the norm in all states, appropriate aspects of evaluation efforts have been specified that are best-suited for the nature of the efficiency program practices that are being implemented there.

Standardized evaluation approaches and protocols

Clearly, for both the US and EU energy efficiency markets, standardized protocols for evaluation are very helpful, and the authors believe there would be merits if such protocols were consistent between the US and EU, and between US states and EU member states. As will be described in this section, there are already numerous well developed evaluation methodology documents that have applicability for both the US and the EU. Having such established and reliable protocols ensures that evaluations produce the necessary results and that those results are developed in an objective and purely unbiased manner. In the authors' opinions, the most effective consequence of the introduction of evaluation protocols in the US to date has been to establish a common language describing levels of rigor and other terminology for buyers and sellers of such evaluation services. This likely sounds like a modest achievement to some of the existing protocols' au-

thors, who perhaps dreamed that they would be declaring new evaluation measurement and verification (M&V) commandments based on their wisdom. Many who are newcomers to the introduction of energy efficiency mandates may not be aware of how mature many aspects of the evaluation practice actually are. While there are areas of evaluation that are still developing and there is always room for new developments, it is surely best to leverage the experience of a past generation.

One goal reached (on the impact side) through the introduction of standardized evaluation protocols and services has been to ensure accuracy as we verify savings through metering, both pre-installation and post-installation, of various energy efficiency measures included in various projects, technologies, and sectors leading to the determination of proper energy baselines and the verification/determination of projected/actual savings achieved (acquired) and to inform future program impact evaluation activity wherever possible.

The benefits of standardized evaluation protocols are categorized in the following sections.

GUIDELINES AND APPROACHES

An effective use of protocols has been to provide guidelines on how to determine which evaluation procedure to apply to a certain situation (type of project, program, sector, etc.) and how to conduct that evaluation. An example of a protocol that has been helpful in this area is the *Uniform Methods Project (US Department of Energy (DOE) Uniform Methods Project for Determining Energy Efficiency Program Savings)*.

TEACHING TOOLS

There also are several documents referenced or titled as "evaluation protocols" that, as a practical matter, are used mostly as reference or teaching tools. A key example of a reference protocol that serves in this capacity is the *California Evaluation Framework (California Public Utilities Commission (CPUC), California Evaluation Framework)*.

Unfortunately, not all protocols are successful in truly benefitting the energy efficiency market. Those protocols that too rigidly specify required procedures can be very limiting. The most aggressive form of protocol – the *rigidly mandated courses of action* – has been and continues to be attempted and reflects the most inflexible definition of the word "protocol," but it has been the least successful. Even when plainly written as such and expected to be used in that way, they rarely work in that fashion. The world of energy efficiency project implementation and the programs that promote them are just too *muddy* – projects are too unique and evaluation funding is too limited.

CUSTOMIZING PROTOCOLS

Many evaluation practitioners use protocols but then customize them to best suit the program and projects being studied, ensuring cost-effective and thorough evaluation M&V under the given financial constraints. The customized protocols describe the process of developing, executing, and reporting on M&V plans. Key steps typically include efficiency project file review, writing and issuing advance letters prior to customer contact, initial customer contact, M&V planning, site visits, analysis, and reporting. Plans are constructed at a level of rigor commensurate with engineering needs and budget and optimized to minimize uncertainty.

Even with such customized evaluation protocols and plans, topics such as the level of rigor for site visits are often based on the building blocks described in a seminal protocol like the *International Performance Measurement & Verification Protocol (IPMVP)*. Such a protocol very explicitly describes the intensity of engineering site rigor. A customized presentation of their data and the estimated level of effort are as follows:

> *Enhanced:* The highest level of engineering rigor. Typically follows the IPMVP* Option B (retrofit isolation, full) or Option D (building or systems modeling). Includes logging equipment performance over time, usually at least 2 to 4 weeks and for as long as a year before and after the modifications. Typically 40 to 160 man-hours each.
>
> *Basic:* Moderate level of rigor. Typically follows the IPMVP Option A (retrofit installation, key parameters only) or Option C (whole facility metering). Usually includes short-term logging. Typically 20 to 60 hours.
>
> *Verification:* Lower level of analytical rigor, typically includes an inspection site visit, possibly spot measurement, and intensive review of the applicant's savings calculations combined with crosschecking against standards and rules of thumb. Typically 12 to 24 hours.
>
> *Desk review:* Similar to verification, but without the site visit. The level of independent analysis can vary and must be explicitly described in advance. Typically 2 to 16 hours.
>
> *International Performance Measurement & Verification Protocol.

An evaluation protocol may require full replication of prior analysis or the completely independent development of savings estimates. It is important to understand that the nature of evaluation is to fully determine project savings and not just depend on the efforts of others. For cost-effectiveness, as previously stated, a sample of project sites, rather than a census, is frequently used.

CAUTION ON CREATING NEW PROTOCOLS

An evaluation consultant may have the impetus to promote the development of a new protocol or standardized approach. Such an effort may enable an individual or firm to present themselves as being at the pinnacle of their profession – the experts among experts, able to tell the rest of the world the way they should do things. And performing such work may give one a sense of doing something important "for the profession." It has similar leadership attraction for potential clients funding similar efforts. However, one must be judicious in launching any new protocol. In the past 3 years, the authors have responded to requests for proposal RFPs or performed work that references over twenty different protocols in the US alone. Rarely in our work have we said to ourselves, "What we are lacking is a new protocol to guide us down the path to the right evaluation plan." To be sure, many protocols have narrow applications. The point is that one should be certain that there is no relevant or closely relevant protocol already available.

LISTING OF MAJOR PROTOCOLS MOST COMMON IN THE US

The following listing is useful in presenting a compendium of already available and proven evaluation protocols in the US. Many or most of these documents will be very helpful in developing a program-specific approach for evaluations.

There is much redundancy between these documents, and many or most of the protocols indeed reference each other. We note that part of the point of providing a listing of all of these documents (and showing links in the reference section) is to point out how easy it may seem to commission a new protocol. Again, we recommend pausing before taking that step as new evaluation protocols should really address novel and/or additional information, such as that described in subsequent sections of this report. It is noted that the listed US-based protocols are placed roughly in order of their traction and maturity in the market; as such, state-specific documents such as those developed in California, have become national documents with common applicability in most states with mature energy efficiency programs.

- *International Performance Measurement & Verification Protocol* (IPMVP) (note that Volume 1 of this seminal document has been produced in fourteen languages).
- *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals.*
- *The California Evaluation Framework.*
- DOE NREL Uniform Methods Project (Department of Energy National Renewable Energy Laboratory).
- ASHRAE Guideline 14 – Measurement of Energy & Demand Savings.
- BPA Measurement & Verification (M&V) Protocol Selection Guide and Example M&V Plan (this is one volume of a 10-volume protocol & reference guide).
- Environmental Protection Agency *Model Energy Efficiency Program Impact Evaluation Guide.*
- DOE EERE Series:
  - Guide for Managing General Program Evaluation Studies.
  - Guide for Conducting Benefit-Cost Evaluation of Realized Impacts of Public R&D Program.
  - Overview of Evaluation Methods for R&D Programs.
  - Impact Evaluation Framework for Technology Deployment Programs.
- Federal Energy Management Program M&V Guidelines: Measurement and Verification for Federal Energy Projects.
- ISO-NE M-MVDR Measurement and Verification of Demand Reduction Value from Demand Resources, (Revision 06) 06-01-14.
- Northeast Energy Efficiency Partnerships (NEEP) Data Collection Protocols and EM&V Forum Glossary of Terms.
- Ontario Power Authority, Canada – Conservation First 2015–2020: Evaluation, Measurement and Verification Protocols and Requirements.

Virtually every US state in which the authors work has standard performance contracting (ESCO) M&V guides and/or evaluation protocols, as well. Examples include the *California Standard Practice Manual* and the *New York Evaluation Plan Guidance for EEPS Program Administrators*.

COMMENTS ON EU EVALUATION APPROACHES

In regard to evaluation efforts in the EU and its member states, the authors acknowledge that they have limited experience with evaluation protocol developments and associated approaches used throughout the EU. We do understand that a full scale EU evaluation guide was considered in the EMEEES project, but that effort was never completed. We also understand that the IEA guides and the EU DSM/EE guides all reference evaluation, and we know that several member states have for years systematically carried out evaluations with approaches customized for their particular program and project needs. We are hopeful that the US-based protocols referenced above will represent a possible framework for future discussions and applicability to the EU energy efficiency and demand response marketplace. We recognize the unique and numerous differences in the US and EU markets, but still believe it would be helpful to initiate and expand a continuing dialog on how best to conduct evaluations in the changing energy efficiency environments.

The quest to improve evaluation

The above listed and described evaluation protocols present a large volume of information on how to conduct a standard impact or process evaluation of energy efficiency projects and programs. With that said, there is a general quest to improve evaluation and a continual push to develop enhanced approaches to improve evaluation response time, quality, and value.

Evaluations are performed for two main reasons. First, they are used for the purpose of reporting on the program and gauging the success of program operators, with information going to both the operators and regulators. Second, information from evaluations can support program quality control and program operator enhancement efforts. All of the detailed findings provide valuable information on the specific strengths and weaknesses and deficiencies of the program, whether on the technical side (with general insights from impact evaluations) or on the procedural side (with insights coming from process evaluations).

Typical evaluation protocol dictates that evaluation activities occur after the completion of a program year, sometimes immediately following that year, but often one or more years after that. Further, evaluations may not occur for every program year; rather, the evaluation efforts may occur periodically to minimize evaluation cost impacts.

EVALUATION EFFORTS SHOULD SUPPORT IMPROVED PROGRAM OPERATIONS

The authors of this paper believe that the time gaps associated with delayed evaluations reduce the benefits that evaluations can provide to program implementers. Results and findings that are delivered one or more years from the date of project and program activities only function as a tardy reporting mechanism. In an effort to enhance evaluations and improve their overall value, we recommend the following:

- **Careful coordination of evaluation with program delivery** – In an effort to ensure that third-party evaluators are

objective and that no bias comes into their work, there is often a strong wall between those entities that handle evaluation and those that deliver programs. Objectivity and the development of accurate results is of primary interest, but the quest for unbiased results should not be done in a manner that limits contact and any level of collaboration between the implementers and evaluators. Some level of common understanding and objectives would better serve the desire to progressively improve efficiency programs and to develop findings that achieve the value of a collaborative process where all parties are seeking to improve efficiency efforts.

- **Integrating impact and process evaluations** – As previously mentioned, impact and process evaluations are often performed independently, with little corroboration of findings between those evaluation teams and often duplication of key data collection efforts. As described in the subsequent section of this paper, there are considerable benefits to be achieved through joint or integrated impact-process evaluation work.
- **Incorporating pre- and post-M&V evaluations** – With impact evaluations that occur months or years later than the delivery dates being assessed, one key piece of information that is left to assumption or to the trust of information in original files is the exact nature and performance characteristics of the pre-installation situation. Generally, that pre-installation condition is poorly documented, so the evaluator must make assumptions to determine the energy use for that portion of their analyses. Alternatively, it is highly desirable to bring evaluators into their role much earlier in the implementation process, allowing for their direct involvement in the assessment and data gathering of the pre-efficiency systems. This recommendation and approach is further developed in a later section of this paper, with specific examples outlining the methodology, and demonstrating enhanced protocols for both project implementation and later evaluations.
- **Real-time evaluation** – A next logical step after the addition of the pre-installation M&V activities is to make the evaluation truly a real-time effort. In this case, immediately following project installation (and prior to installation, too, if at all possible), M&V systems are deployed and the results for those projects can be made available to evaluation management, project and program implementers, and regulators. Further, aggregated program results can be progressively updated, providing overall results to all parties. This area is also discussed later in this paper.

Integration of impact and process evaluations

Process and impact evaluations have in the US traditionally been conducted separately due to differences in implementation schedules, funding, and objectives. Regulatory reporting requirements drive primary evaluation objectives, such as determining savings and realization rate, yet there is also an expectation that the evaluation findings will result in actionable recommendations for improving program design.

The process team focuses on the big picture and the scope can include assessment of program delivery, marketing and outreach, incentive levels and structures, benchmarking, and customer (end-user) satisfaction. Assessment of program delivery is focused on whether the design of the program is appropriate. Acting in parallel, the impact team is more narrowly focused on producing a single number: the net program savings. To arrive at this number typically requires disciplined in-field data collection. While the impact team engineers use the applications and project files to supplement their data collection, they rarely assess how those conform to program design intent.

As a result of their independent objectives, the process and impact teams make independent recommendations, which may be contradictory. For example, a large commercial/industrial efficiency program recently underwent separate, concurrent process and impact evaluations. The process evaluation team assessed the program's tracking system design and quality checks and concluded that the design and quality control steps were adequate. Concurrently, the impact evaluation team closely tracked savings claims at a site-specific level over the course of on-site M&V for a large number of projects. Contradictory to process team findings, the impact team concluded that the program's tracking system had major errors, such as duplicate records and a high incidence of failure to update records when applications were revised. Although both teams had useful observations, they were flawed in that neither of the recommendations provided program implementers with a coherent direction.

The reason for integrating process and impact efforts is to provide more powerful and actionable recommendations for program improvements. For recommendations to be compelling, the implementers must be provided with evidence, an action to undertake, and also a yardstick for assessing the value and the cost of implementation. By joining forces, recommendations to improve the program can better meet these standards.

While our focus is on the improved recommendations, the overall benefits of an integrated approach would include:

- **Lower management costs** – Integration and economies of scale will reduce redundant planning, data collection, and reporting efforts, allowing the evaluation team to do more with the same budget.
- **Increased technical quality** – Impact evaluations answer the question of how much energy savings occurred as a result of program efforts. Process evaluations answer the question of how effectively a program is working. Integrating the evaluations together improves evaluators' understanding of why a program achieved the savings levels it did.
- **Faster service** – Familiarization activities that include both impact and process team leaders at the same time (or they can be the same person), reduce the calendar time requirements. Joint survey instruments will not need to cycle back and forth between evaluation teams, again saving time. In cases in the US where moves have been made to integrate process and impact, program managers have claimed considerable new benefit.
- **Reduced customer and implementer burden** – There will be fewer interviews with staff, allies, and participants.
- **Recommendations with a price tag** – The integrated approach will allow quantified improvements in savings estimates to be coupled with the elements of the cost to implement key recommendations.

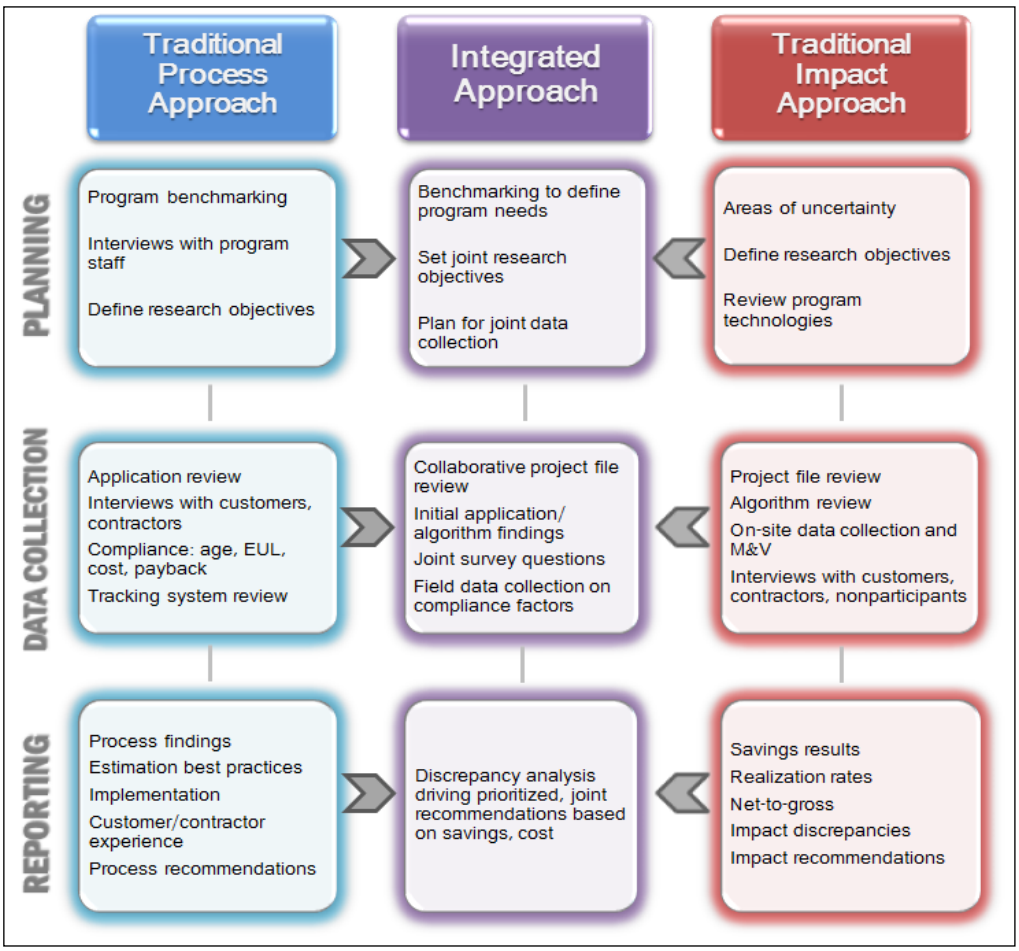

Figure 1. Benefits of an integrated process/impact approach at various evaluation phases.

The authors believe this approach elevates the "actionability" of the recommendations. Process evaluations often "fly blind" because the budgets are not sufficient to gather in-field data that could more precisely inform their findings. Armed with robust site-level detail, the final program recommendations can identify the impact on savings due to these errors and identify corrective actions that can be used to determine cost.

The result is process evaluation outcomes that are enhanced by engineering and an impact evaluation approach that is informed by process conditions on the ground. The practical procedures of both evaluations and how they link together are highlighted in Figure 1.

Key components of an integrated impact-process evaluation, along with some major benefits, include:

- **Integrated planning** – The master plan for an integrated evaluation will identify joint areas of investigation, synergistic data collection, and strategies for developing joint findings and recommendations. This will save cost and enable comprehensive and complementary research.
- **Integrated data collection** – Integrated data collection enhances the quality of both evaluation branches in a way that is not possible with their separation. Specific examples of areas where we see opportunities to increase quality through integrated data collection include the following:
  - Examining savings algorithms and data collection techniques help to determine alignment with program processes.
  - The combined approach facilitates root-cause analysis of problems related to measure ineligibility and improper baseline selection.
  - The integrated approach improves the ability to quantify sources of discrepancy and then tie recommendations to actual expected outcomes.
  - Integrating data collection by incorporating two surveys into one will save money by eliminating the one-time costs that are inevitably associated with each standalone survey, such as computer-assisted telephone interviewing centre setup.
- **Integrated recommendations** – The integrated approach will result in coordinated, consistent, and possibly unified recommendations. One of the most valuable outcomes of the integrated approach is that recommendations will be grounded in actual program outcomes – savings and realization rates. This is the key ambition of an integrated evaluation process and is not always a natural basis of findings in traditional evaluations. By offering shared conclusions and recommendations that come, where appropriate, with esti-

mated savings and costs, the process evaluation will become accountable to the impact evaluation.

Ultimately, integrating recommendations could make it possible to create a managerial tool that would result in magnitude-level estimates of the change in savings or program costs that would result from implementing each recommendation. Steps to do this would include:

1. High-resolution discrepancy analysis, which links realization-rate gaps to specific technical and process factors that differed between the implementer and evaluator analyses. This analysis can be directly translated to kWh or therm impacts, as has been shown in countless impact evaluations conducted throughout the US.
2. Project applications analysis including a systematic desk review of project-specific documents to evaluate application quality to link administrative and process weaknesses to impact outcomes. Overarching assessments are typically a product of process evaluation; project-level review is typically a part of impact evaluation. Project applications analysis will link these two together.
3. Savings algorithm review and assessment, which will identify technical process gaps that leave savings on the table and allow clear quantification of their impacts.
4. Association of process evaluation recommendations with specific quantitative consequences to program impact. While the estimates may be coarse, the effort will greatly enhance the overall picture.
5. In-field assessment of lost opportunities.

The purpose of these components is to link more general impact findings on net-to-gross and technical realization rates with action-oriented, program-component-specific process recommendations. Similarly, recommendations from project-specific impact work will be rolled up into the same best practices process evaluation framework allowing program managers to receive *one* set of organized, actionable recommendations that are connected to actual, quantifiable program outcomes.

Incorporating pre- and post-M&V

It has been clearly demonstrated that dynamic integration of evaluators into key phases of the project installation, enabling pre-M&V will facilitate better savings estimates and ultimately, improved programs. As the typical impact evaluation process currently stands, program managers must wait for impact evaluation results that may take years to properly identify the "true" ex-post savings associated with certain measures and projects. Alternative M&V initiatives enable the metering of various pre- and post-installation energy efficiency measures included in various projects, technologies, and sectors leading to the determination of proper energy baselines and the verification/determination of projected/actual savings achieved (acquired). Further, this pre/post M&V approach can also focus on the determination of savings associated with new or customized energy efficiency technologies.

DEFINING BASELINES AND UNDERSTANDING THE PRE-INSTALLATION OPERATIONAL STATUS OF EQUIPMENT

Determination of energy efficient measure savings is not just dependent on the performance and operating characteristics of the newly installed systems, but also on the existing or baseline technologies and operation. A fundamental benefit of the pre-installation M&V process is that it enables the best possible insights into that base case, ensuring that savings are accurate, assumptions are consistent, and project achievements are neither over- or understated.

Since evaluation-based pre-installation metering must so closely integrate with the activities of program implementers, there are several concepts that must be carefully addressed in planning this type of evaluation:

- **Establish dynamic communications with both evaluation and program managers** – The pre/post M&V approach is occurring dynamically within the exact time frame of the overall project installation, and since it will remain necessary to perform the evaluation activities with a third-party objectivity, it will be essential to establish a dynamic working relationship with both evaluation contract managers and program/project delivery managers. Efforts must be made to interact with the project development process, but avoid influencing it inappropriately.
- **Develop comprehensive evaluation M&V plan for each project assessed** – To ensure that there is a predictable path of evaluation activities, it will be necessary to develop a site M&V plan for each project reviewed (typically following procedures as specified in the IPMVP protocol document) and detailing the best IPMVP option to be used. As with a post-installation evaluation site visit, a comprehensive plan will be developed for each pre-install M&V site visit. This will address details of the facility, measures to be investigated, file or document review approach, pre- and post-M&V approaches, metering systems to be deployed, data to be collected, and analysis methodologies to be used.
- **Intercede with engineering support if necessary and appropriate** – While all of the specified pre-installation evaluation activities, data collection, and assessment efforts are intended to be done in an independent and objective manner, without undue influence on the project installation team, there is a common objective that this dynamic pre-installation evaluation effort provides clear value to the efficiency process. Thus, where projects are demonstrating clear weaknesses or issues or deficiencies, as part of the dynamic evaluation process, an approach must be defined for delivery of rapid evaluation-issue findings for which evaluation managers can communicate with program staff for actions on appropriate corrections. Any such actions should be clearly documented.
- **Avoid over-interaction with implementation market actors** – Regardless of the above point, evaluation objectivity must be intact. Therefore, evaluation staff should minimize site conflicts and avoid biasing project plans, and any planning with installation staff, contractors, and equipment vendors should be kept to a level that does not end up changing installation actions unless approved by the overall team.

- **Limit metering must to short-term data logger deployment** – It must be noted that the desire to deploy metering and data logging equipment on the existing equipment, prior to efficient equipment installation, will represent some level of intrusion into the project. Such intrusions must be minimized and data logging must be kept to the minimum necessary time frame to reduce the impact on installation schedules.
- **Major problems and findings from the pre-installation M&V must be addressed immediately** – As with previous points, any initial problems from pre-installation M&V must be addressed immediately as part of the evaluation process. A benefit of this dynamic evaluation process is that problems can be readily addressed rather than just reported on retrospectively.
- **Integrate with post-installation M&V to develop evaluation results** – Once the project has been fully installed and is in operation, the post-installation M&V process should be initiated. Prompt initiation allows for an optimized comparison with the base case, reducing the chance of results being influenced by overall changes in system operation caused by other factors. Metering equipment deployed post-installation should be consistent with the pre-installation approaches, enabling optimized investigation of system energy use, savings, and operating characteristics. Detailed analyses will determine the new system's energy use and resulting gross savings.
- **Gather key process information** – Gathering ancillary process information during both pre- and post-installation site visits and M&V will provide more timely and more accurate information on project decision-making criteria, issues with the installation, details on operational characteristics and anticipated changes, and potential insights and data non-energy benefits.

Regardless of efficiency program structure, independent pre/post evaluation and M&V strategies outlined here have great value in swiftly supporting and developing higher levels of program quality, dynamically facilitating delivery of performance information that can guide future projects.

One key example is the pre-/post-M&V evaluation efforts being conducted in New York State for Consolidated Edison Company of New York (CECONY). In two jurisdictions, CECONY has been given mandates to swiftly install efficiency projects to resolve major problems associated with load growth, T&D infrastructure limitations, and closing power generation facilities. In the efficiency projects being installed through this mandate, ensured performance and savings of the installed equipment is absolutely necessary to provide confidence that more costly infrastructure upgrades can be avoided. As such, CECONY has adopted a rigorous pre/post evaluation process that immediately demonstrates project savings and overall program results. Absent the pre-installation metering component and the swift deployment of post-installation metering, savings confidence would be dramatically reduced and evaluation results would be considerably delayed. To date, this continuing effort has provided instrumental value to CECONY and New York State.

Benefits of real-time evaluation

For many energy efficiency and demand management programs, determination of immediate results can support the effort toward strategic planning and monitoring for infrastructure planning and emission reduction and mitigation. Impact and process evaluations on the traditional time scales (i.e., after the program cycle is complete) are not particularly useful to the directors of efficiency efforts or to the planners who need to determine whether programs will have procured the energy efficiency impacts and demand reductions as specified in various directives. A post-program impact evaluation will not alert planners in time, and a post-program process evaluation will not provide actionable findings that immediately improve the program operation.

In the end, the most useful output of evaluations will be real-time feedback to assess the actual progress toward the goals and provide suggestions to make the program more effective and progressive. Such a real-time evaluation exercise readily flows from the more dynamic pre/post evaluation protocol discussed in the prior section of this paper. Taking the next steps, the real-time evaluation would have the following characteristics:

- **Integrated real-time evaluations** – Real-time evaluations will build upon program pre/post M&V activities described above, incorporating both program impacts and processes. Such efforts will decrease duplicative effort and will be more cost-effective overall than separate post-program evaluations, providing near immediate results to project and program implementers, evaluation managers, and regulators.
- **Dynamic reporting of site, project, and overall program impacts** – The real-time evaluation will provide on-going data to the program staff and regulators to support reporting and planning. This timely information will be valuable in targeting program efforts to the most productive measures and customer types and can help infrastructure planning teams with up-to-date information for forecasting, as well as energy-focused efficiency reductions.
- **Assured quality control and immediate programmatic and project improvements** – Through dynamic reporting of pre- and post-M&V of newly installed projects, much data can be ascertained on the merits of a wide variety of technologies, approaches for implementation in various sectors, and assumptions on operating characteristics that drive technology success. This information can be used to guide future projects – even ones that are near ready for implementation planning – in enabling better installations that are more likely to achieve their projected savings. In addition, programs can learn from the immediate insights, enabling improved targeting of appropriate sectors and technologies for the most significant and successful program and project impacts.
- **Real-time dashboard and database of program results** – The mechanisms for real-time reporting and information dissemination are crucial for the success of real-time evaluation efforts. Typical strategies include graphic reporting user interfaces (on-screen dashboards), with associated backbone databases, that are continually updated to reflect results on specific projects and aggregated results. The au-

thors have developed and applied such structures in some on-going real-time evaluation efforts to great success. Recent efforts for CECONY, National Grid, and NYSERDA have proven highly successful and being expanded to capture more details of the installed efficiency projects.

- **Progressive move toward incorporation of real-time metering with immediate reporting of impacts** – The ultimate stage in real-time evaluations involve use of actual, real-time metering system reporting, with cloud-based meter result distribution and storage and ultimate analysis and aggregation to stated database structures and reporting dashboards.

Conclusion and discussion

This paper has described the needs for technically sound and defensible evaluation approaches and has described the key protocol documents that have gained traction in the US energy efficiency marketplace. Such protocols have been developed over the course of many years of experience in evaluation in the US. Most of the cited protocols could have great applicability for both the US and EU markets, despite their very different market characteristics.

The authors have further discussed the most recent and next evolutions in evaluation, including integrated impact and process evaluations, pre-and post-M&V evaluations, real-time data collection, analysis, and reporting of evaluation research findings. These new approaches should be considered for incorporation in the next generations of existing evaluation protocols and could be readily adopted for improved and consistent evaluations in the US and EU.

In general, the adoption of standardized evaluation approaches is inconsistent between US states and EU states, and there are generally different approaches used in the US and EU. While there may be aspects of the US and EU markets that justify major differences in strategies for evaluation, the authors of this paper still believe it would be valuable to at least consider higher levels of US-EU evaluation planning and collaboration. As such, we suggest convening a leadership panel of US and EU program managers to discuss commonalities and unique differences between the US and EU markets and outline where evaluation approaches should be different, and where there should by synergies and common methods. In cases for which it may be possible to collaborate on evaluation protocols and strategies, such an expert panel should meet with appropriate regularity to ensure that the cost-effectiveness and benefits of collaboration are achieved.

In both the US and the EU, working together or independently, increasing consideration there must be given to new M&V technologies that can make evaluation more streamlined and cost-effective. Further, several new approaches such as those that address integration of impact and process evaluations deserve further consideration and potential incorporation into existing evaluation protocols. Again, approaches that can enhance development of useful data, improve reporting of program impacts, and aid in simplifying the quest for higher demand side efficiencies should be targeted, assessed, and brought into standard evaluation approaches.

Finally, the evaluation of energy projects and programs must evolve and become fully immersed in the changing technology paradigm, where the evolution of remote monitoring and real-time metering systems are becoming increasingly evident. Such developments can and will become the mainstays of project operational monitoring, and it is increasingly likely that all projects, large and small, will be installed with monitors that will cost-effectively and dynamically report on systems' performance. Clearly, such developments will necessitate the need for changing yet consistent evaluation approaches that can make appropriate use of newly available data. With such approaches taken, programs and projects will improve as evaluators report more accurately on their actual energy and demand achievements.

References and links

American Society of Heating, Refrigeration, and Air Conditioning Engineers (ASHRAE) Guideline 14, Measurement of Energy and Demand Savings, 2002. https://gaia.lbl.gov/people/ryin/public/Ashrae_guideline14-2002_Measurement%20of%20Energy%20and%20Demand%20Saving%20.pdf.

ANSI Energy Efficiency Standardization Coordination Collaborative (EESCC) Roadmap Chapter Four: Evaluation, Measurement, and Verification (EM&V), 2006. http://www.ansi.org/standards_activities/standards_boards_panels/eescc/FINAL_EESCC_ROADMAP_ONLINE.pdf.

Bonneville Power Administration (BPA), Measurement & Verification (M&V) Protocol Selection Guide and Sample M&V Plan, 2012. http://www.bpa.gov/EE/Policy/IManual/Documents/July%20documents/1_BPA_MV_Protocol_Selection_Guide_May2012_FINAL.pdf.

California Public Utilities Commission (CPUC), California Energy Efficiency Evaluation Protocols, 2006.http://www.calmac.org/events/evaluatorsprotocols_final_adoptedviaruling_06-19-2006.pdf.

California Public Utilities Commission (CPUC), California Evaluation Framework, 2004.http://www.calmac.org/publications/california_evaluation_framework_june_2004.pdf.

California Public Utilities Commission (CPUC), California Standard Practice Manual: Economic Analysis of Demand-Side Projects and Programs, 2001. http://www.cpuc.ca.gov/NR/rdonlyres/004ABF9D-027C-4BE1-9AE1-CE56ADF8DADC/0/CPUC_STANDARD_PRACTICE_MANUAL.pdf.

Directive of 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on Energy Efficiency, 2012. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012L0027&from=EN

EMEEES: Evaluation and Monitoring for the EU Directive on Energy End-Use Efficiency and Energy Services, April 2009. http://www.evaluate-energy-savings.eu/emeees/downloads/EMEEES_ProjectDescription_completed.pdf.

Haeri, Hossein and Tina Jayaweera, *EM&V Methods: A Time for Uniformity,* 2014 International Energy Policy & Program Evaluation Conference, Berlin. http://www.iepec.org/conf-docs/papers/2014/Hossein%20Haeri.pdf .

Hewlett, Patrick and Sue Haselhorst, *Merge Ahead: Integrating Process and Impact Evaluations*, ACEEE 2014.

International Performance Measurement and Verification Protocol (EVO IPMVP), 2003-2014. http://www.evo-world.org/index.php?lang=en.

ISO New England Manual for Measurement and Verification of Demand Reduction Value from Demand Resources, 2014. http://www.iso-ne.com/participate/rules-procedures/manuals.

New York Department of Public Service (NY DPS), Evaluation Plan Guide for EEPS Program Administrators, 2013:http://www3.dps.ny.gov/W/PSCWeb.nsf/96f0fec0b45a3c6485257688006a701a/766a83dce56eca35852576da006d79a7/$FILE/NY_Eval_Guidance_Aug_2013.pdf.

Northeast Energy Efficiency Partnerships (NEEP) Regional Evaluation, Verification and Measurement Forum, Data Collection Protocols, 2012. http://www.neep.org/data-collection-protocols-0.

Stern, Frank and Dimitris Vantris, *Protocols for Evaluating Efficiency – Both Sides of the Atlantic,* 2014 International Energy Policy & Program Evaluation Conference, Berlin. http://www.iepec.org/conf-docs/papers/2014/Frank%20Stern.pdf.

US Department of Energy (DOE) Uniform Methods Project for Determining Energy Efficiency Program Savings, April 2013. http://energy.gov/sites/prod/files/2013/07/f2/53827_complete.pdf.

US Department of Energy Federal Energy Management Program (FEMP), 2008.http://www1.eere.energy.gov/femp/pdfs/mv_guidelines.pdf.

US Department of Energy Office of Energy Efficiency and Renewable Energy (DOE EERE), EERE Guide for Managing General Program Evaluation Studies, 2006. http://betterevaluation.org/sites/default/files/evl_mg_app.pdf.

US Department of Energy Office of Energy Efficiency and Renewable Energy (DOE EERE), Impact Evaluation Framework for Technology Deployment Programs, 2007. http://www1.eere.energy.gov/analysis/pe_index.html.

US Environmental Protection Agency (EPA), Model Energy Efficiency Program Impact Evaluation Guide, 2007. http://www.epa.gov/cleanenergy/documents/suca/evaluation_guide.pdf.