Evaluation and Regulation: Put to the Test

William P. Saxonis, New York State Department of Public Service, Albany NY[1]

ABSTRACT

In 2008, New York State adopted one of the most ambitious energy efficiency programs in the nation, the Energy Efficiency Portfolio Standard (EEPS). In approving a portfolio of approximately 100 EEPS programs, the New York State Public Service Commission stressed the importance of rigorous program evaluation and took several concrete steps to provide evaluators with the necessary tools to achieve this objective. The Commission's policies created a new evaluation paradigm that has now been in place for over two years. This is an opportune time to analyze New York's record in meeting challenges in several key areas including finding the "correct" level of regulatory oversight relative to the evaluation process, managing the various and sometimes conflicting demands and expectations of multiple stakeholders, and adopting statewide and regional approaches to evaluation. Like energy efficiency programs, the policies and procedures used in the program evaluation process must also be regularly evaluated. An examination of evaluation efforts nationally makes clear that simply spending more money and introducing new polices does not automatically translate into success. This paper provides eleven "lessons learned" from the New York experience that will have wide applicability to evaluators everywhere.

Introduction

For over three decades, New York State has been a national leader in energy efficiency initiatives, recognizing the economic, environmental and social benefits that energy efficiency delivers to its over 19 million residents. In recent years, we have seen a dramatic increase in investment in energy efficiency programs, not only in New York, but nationally. Between 2007 and 2010, investment in electric efficiency programs in the United States doubled from $2.7 billion to $5.4 billion, and expenditures for natural gas efficiency experienced an even greater percentage increase (CEE, 2011). In June 2008, the New York State Public Service Commission (Commission) approved an Energy Efficiency Portfolio Standard (EEPS) with the goal of reducing electricity usage by 15% from the level forecasted for 2015, and, in May 2009, established a comparable efficiency goal for natural gas (Case 07-M-0548). By January 2010, the Commission had approved about 100 new programs, increased the role of utilities in program administration, and more than tripled annual expenditures for energy efficiency. Total energy efficiency program spending, under the jurisdiction of the Commission, now exceeds $600 million annually, an increase of approximately 240 percent over the 2007 funding level. Other major contributors to the EEPS energy saving goals, but not under Commission jurisdiction, include programs offered by the New York Power Authority (NYPA) and the Long Island Power Authority (LIPA) and the effect of state building codes and appliance standards. The increasingly significant role of energy efficiency in New York's energy future has elevated the need for reliable evaluation results that will document past achievements, provide insights to guide future programs and

[1] This paper reflects the views of the author and does not necessarily, implicitly or explicitly, express the views of the New York Public Service Commission or the Department of Public Service. The author acknowledges the assistance of his evaluation team, Katie Mammen and Donna DeVito, in the preparation of this paper.

policies, and determine if New York's energy programs are returning fair value to ratepayers. The evaluation results will answer fundamental questions, such as:

- Are the EEPS goals being met?
- Can the New York Independent System Operator (NYISO) rely on the estimates of energy efficiency program impacts to meet their statewide planning and forecasting needs?
- Will estimates of energy efficiency impacts be sufficiently accurate to allow equitable lost revenue recovery and incentive payments to the utilities?
- Are the right programs being offered at the right times?
- Are the programs being implemented effectively?

As the Commission developed the EEPS, it also increased its commitment to rigorous, timely, and transparent program evaluation and created a new evaluation paradigm.[2] This enhanced commitment to evaluation was not intended as a criticism of previous evaluation efforts, but as a recognition that as investment levels increase, so does the need for more rigorous evaluation that can reliably assess program design and performance. The Commission's actions were based on the premise that you cannot accurately evaluate energy efficiency programs without having fundamental policies and procedures in place. Poor planning, sloppy data collection and inadequate resources can doom any evaluation and defeat the best intentions of regulators.

Past Commission actions designed to provide the tools to transform these high expectations into reality include:

- Increasing funding for evaluation by 150%
- Establishing evaluation guidelines and standards
- Forming the Evaluation Advisory Group (EAG), which represents about 25 interested parties, to advise the Commission and Staff[3] on evaluation issues
- Endorsing a more active role for Staff in overseeing and guiding evaluation and reporting activities.

A previous paper for this conference (Saxonis, 2009) addressed the then new and evolving evaluation paradigm. The paper reported that "New York's energy efficiency community is working diligently to develop a new paradigm to provide the foundation to evaluate a new generation of EEPS energy efficiency programs." This new evaluation paradigm has now been in place for over two years, and it is an opportune time to analyze New York's record in meeting challenges in several key areas, including finding the "correct" level of regulatory oversight relative to the evaluation process, managing the various and sometimes conflicting demands and expectations of multiple stakeholders, and adopting statewide and regional approaches to evaluation. Like energy efficiency programs, the policies and procedures used in the program evaluation process must also be regularly evaluated. An examination of evaluation efforts in other states makes clear that simply spending more money and introducing new policies does not automatically translate into success. Written from a regulatory perspective, this paper offers "lessons learned" from New York's experience that will be of value to the evaluation community at large.

---

[2] More information about evaluation in New York can be found at the Department of Public Service web page. http://www.dps.state.ny.us/EEPS_Evaluation.html General information about EEPS: http://www3.dps.state.ny.us/W/PSCWeb.nsf/All/06F2FEE55575BD8A852576E4006F9AF7?OpenDocument

[3] In this paper, Staff refers to the staff of the New York State Department of Public Service.

Evaluation Guidelines

During the first year of the EEPS, Staff's evaluation effort focused on building a solid foundation to meet the evaluation challenges of the expanding size and scope of the EEPS program portfolio. While there is always a temptation to get side tracked by the latest state-of-the-art evaluation techniques and the current "hot topics" in evaluation, the reality is that to effectively tackle the challenge of evaluating program impacts it is essential to put evaluation fundamentals in place first. Even the most advanced and proven methodologies will fail unless they are fueled by reliable and unbiased data. Accordingly, Staff began by encouraging an open discussion of evaluation issues among the key parties (e.g., program administrators, evaluation experts) and establishing clear evaluation guidelines, including a roadmap for collecting the data necessary to conduct first-rate analysis.

This "back to basics" philosophy was driven by past experience. For example, in 1993 independent consultants retained by the New York Power Pool reported that "PSC staff, as well as a number of utility staff, expressed frustration in trying to compare the results of similar programs among the utilities. The data was inconsistent in completeness, format and definition." (Barakat & Chamberlin, 1993). Moreover, many evaluations in New York and elsewhere have been compromised by easily avoided problems arising from deficiencies in basic evaluation planning and design, such as a failure to collect the necessary data (e.g., customer type, total cost of measures).

An important first step was to establish Evaluation Guidelines. As outlined in a June 2008 Commission order initiating the first round of EEPS programs, Staff was required to issue Evaluation Guidelines within 45 days with the objective of putting evaluation standards in place and offering direction for producing quality evaluation and oversight for the complete range of EEPS programs.[4] The resulting Evaluation Guidelines addressed key areas such as evaluation methodology (e.g., statistical rigor, process and impact evaluation); ethical and operational standards; data collection policy; and reporting. Anticipating the diversity of the EEPS program portfolio, Staff rejected a "one size fits all" evaluation approach and crafted guidelines that balanced tough standards with the flexibility to permit evaluators to use the most reliable, practical and cost-effective methodologies. For example, Staff set the accuracy target for customer surveys and estimating program energy savings at the 90/10-confidence/precision level. At this level, one can be 90 percent confident that the measured value (e.g., the energy reduction resulting from a program) is within +/- 10 percent of the reported value based on sampling techniques. Staff recognizes that this is a rigorous standard and depending on the program type and population size, may prove impractical or too costly to achieve. Moreover, the value of a rigorous precision and confidence level is dramatically reduced if the sample is significantly biased. Therefore, Staff allows flexibility regarding confidence and precision, but only when the evaluation plan offers a clear explanation and justification for an alternative standard.

Staff's policy is not to depart from the standards set forth in the Evaluation Guidelines lightly. The case for a waiver must be convincing and the alternative standard must be in the best interest of the ratepayer. Ultimately, Staff did not want program administrators to invest a dollar to document if a program saved fifty cents. When making an assessment of the necessary rigor, Staff also assesses the value of the data in the context of its overall importance to the evaluation effort. The Evaluation Guidelines noted that adherence to the highest evaluation standards and the greatest frequency of evaluation would typically apply to those programs:

- Providing expensive infrastructure investments
- Eligible for utility incentive payments or lost revenue recovery
- Targeted for a significant budget increase

[4] The Evaluation Guidelines are available on the PSC web page: http://www.dps.state.ny.us/Evaluation_Guidelines.pdf

- Producing results far above or below expectations
- Implemented as an innovative program on a pilot basis
- Containing measures with high savings variability
- Based on a limited existing knowledge base
- Making a large contribution to the overall portfolio savings.

Another key pillar of the Evaluation Guidelines is that every proposed program be accompanied by a detailed evaluation plan for Staff review and approval. Developing an initial evaluation plan in preparation for launching a program allows evaluators to work with program planners to identify data collection needs, establish the evaluation approach and synchronize evaluation goals with the program's performance goals. While there are many elements in an evaluation plan, the most prominent components are the process and impact evaluation strategies. Process evaluations are used primarily to assess program design and implementation and also to identify opportunities for program improvement and to track program progress. Impact evaluations are used primarily to quantify energy and demand savings and other potential program impacts such as environmental benefits.

Lesson Learned-Evaluation Guidelines are an Essential First Step

The Evaluation Guidelines have proven valuable especially considering that New York's ten utility program administrators had not operated and evaluated a major portfolio of energy efficiency programs in New York State for approximately 10 years. Furthermore, the New York State Research and Development Authority (NYSERDA), the administrator of New York's System Benefits Charge (SBC) program since 1998 and also subject to scrutiny by the Commission, received a major increase in evaluation funding with the expectation of employing more advanced analytical techniques. Generally, for all the EEPS program proposals and evaluation plans, there has been a strong commitment by the program administrators to adhere to the Evaluation Guidelines. While the Evaluation Guidelines do not offer a detailed roadmap for evaluating every possible program type, they do provide the basics and recognize that methodological detail is more appropriately addressed within the scope of the program evaluation plans. The Evaluation Guidelines were not created in an "ivory tower," and handed out to the program administrators. They were developed with the active engagement of the EAG and Staff, as described in more detail below. This approach allowed Staff to craft guidelines with a general consensus from the EAG that the guidelines were clear, reasonable and likely to be a successful match for the New York's energy program portfolio. In addition, this team approach increased "buy in" for the Evaluation Guideline's objectives. The Evaluation Guidelines were recently updated with the modifications limited primarily to minor refinements.

The Evaluation Advisory Group: A More Detailed Look

The EAG was established in 2008 to serve as a vehicle for encouraging communication and cooperation among program administrators, Staff, and other interested parties on critical evaluation issues. The EAG includes several organizations active in energy efficiency programs that are not under the Commission's jurisdiction. For example, the current EAG membership includes NYPA, LIPA, NYISO, several state agencies, the City of New York and many environmental and business related groups.

Lesson Learned-An Advisory Group is Important but Requires Support

Despite the wealth of experience possessed by EAG members, including several nationally recognized evaluation experts, evaluation is a complex subject requiring knowledge from numerous

disciplines, including statistics, engineering, economics, and research methods (e.g., questionnaire design) and typically requires significant time for analyzing results. Because evaluation is as much an art as it is a science, evaluators' judgment calls frequently generate lively debate. While most sides of the debate reflect reasonable positions, there is usually no obvious and easy answer.

Staff has observed that other advisory groups frequently lack the resources to analyze complex issues effectively. Early on, a major challenge was to determine how Staff could engage the support necessary to allow its internal evaluation team and the EAG to effectively analyze the crucial technical details that could spell success or failure for an evaluation. In addition, Staff knew that it would be enormously valuable to have access to, and an understanding of, the latest evaluation data from outside New York. This insight could help Staff avoid the mistakes of others and alert Staff to the best available research methods. Unfortunately, the sheer volume of evaluation data makes it difficult for regulatory and utility staff to find the time to analyze these often voluminous reports. States like California, Wisconsin, and Massachusetts release thousands of pages of evaluation reports in a typical year. The solution was to engage an independent evaluation contractor to serve in an advisory and research capacity to Staff and the EAG. Staff secured an evaluation firm with experience in multiple states and with multiple program types to provide an additional breadth of knowledge to the analysis. This independent third-party feedback has proven invaluable. Limitations in hiring additional Staff, or even replacing departing staff, also made the use of outside assistance important.

Lesson Learned-Provide Leadership to the Advisory Group

With so many diverse organizations (i.e., varying in perspectives, resources and expertise) represented on the EAG, it became clear that the group should be chaired by one independent organization such as Staff. As a result, Staff organized meetings, offered general direction, provided essential resources (e.g., an internal web portal, meeting space) and assumed responsibility for administrative details. If leadership is too diffuse, effectiveness can suffer. We have seen this happen when Staff experimented with a less centralized approach with EAG subcommittees and observed this problem with other advisory groups. The objective, however, is not to reduce member input, but rather to provide the ingredients for the more efficient operation of the group.

Lesson Learned-Communication (Internal, External) is Essential to Success

In general, the EAG meets once a month, and its subcommittees, which focus on specialized topics and tasks, usually meet more frequently. Sometimes the meetings are in-person in Albany, but are usually conducted as teleconferences. While there is a desire to have more in-person meetings, travel is an issue both in terms of time and cost. Many key members of the EAG are located 150 miles or more from Albany, requiring an investment of five or more hours of travel time for what is often a meeting of about two hours. Staff's ability to travel is limited due to budgetary constraints. Moreover, it is important to respect people's time, especially in this era where interest in energy efficiency is rising and placing an increasingly heavier demand on evaluators, regulators and program implementers. Despite well planned meetings (e.g., detailed agendas are provided several days before the meeting with topics prioritized by time), there is consensus that our meetings would be more effective if we had more face to face contact. Staff is dedicated to finding the appropriate balance of in-person meetings and teleconferences, but this is still a work in progress. Perfecting this type of administrative detail is more critical than it might initially appear because the EAG's role as advisor and problem solver requires effective communication both within the group and to outside stakeholders. For example, an issue recently arose that would have had a chilling effect on the ability of New York's program administrators to conduct process evaluation surveys. A provision of the New York Public Service Law (Article 4, Section 65) states that "No gas corporation or electric corporation shall sell or offer for sale any list of

names of its customers." Staff's legal counsel interpreted "sell or offer for sale" as covering any method of transferring the data, including supplying the information at no charge. Simply stated, a utility could not provide its evaluation contractor with a list of customers not participating in EEPS programs for the purpose of conducting an evaluation-related survey. This issue became a prime topic for the EAG. While utilities could have filed a petition with the Commission or informally contacted various Commission staff with a plea for help, it was far more effective to have a ready-made forum to discuss the law and its consequences. Staff was able to bring their legal counsel to the regular monthly EAG meeting, not only to provide a detailed discussion of the legal interpretation, but also for legal counsel to hear firsthand the potential negative impact on the evaluation process. It was also a chance for everyone involved to understand the passion on both sides of the debate; the consumer advocate's passion to protect consumer privacy and the passion of the evaluators to produce the most useful and reliable results. This exchange eventually resulted in utilities being allowed, under certain conditions, to share contact information with their evaluation contractors.

Lesson Learned-Keep the Advisory Group Structure Simple

In the formation of the EAG, there was much discussion about the structure of the group. For example, should there be a formal voting structure and, if so, should the vote of large utilities count the same as small utilities? Would consensus be necessary on all issues? Staff opted for a more informal structure since the Commission established the EAG's role as advisory. Ultimately, it is the Commission or Staff that makes the final decisions. Generally, EAG members reach consensus. In some cases, however, the EAG might endorse a policy without full agreement, but note any areas of significant disagreement. This approach has allowed the EAG members to arrive at decisions more quickly and, as several EAG members have noted, work together with a consistently high level of cooperation and congeniality within the group. Of course, the less time spent quarreling and settling disputes, the more time the EAG has to focus on important evaluation issues.

Lesson Learned-Keep the Advisory Group Focused on Big Issues, Use Subcommittees for Detailed Technical Issues

The EAG relies on subcommittees to refine, research and report on technical and policy matters. The subcommittees report their findings to the full EAG for review and further direction. One important achievement in this area was reviewing and providing feedback to refine and update New York's Technical Manual. This manual provides uniform, measure-specific approaches to estimate the energy and demand savings achieved by EEPS programs to avoid conflicting energy saving estimates for the same measures within different programs. This effort was the product of several months of work, including reviewing hundreds of comments and recommendations, which resulted in a proposal to update and consolidate the Commission's five sector-specific technical manuals. [5]

The subcommittee approach allows the most knowledgeable and interested people to focus on a well-defined topic area. In addition, technical experts associated with EAG members are allowed to participate in these subcommittees on an as-needed basis. This approach helps prevent the EAG from getting bogged down in technical details that might distract from its primary focus and more effectively allocates resources. The monthly EAG meetings include updates on subcommittee projects, and, when a subcommittee has significant results to report, its members will formally address the EAG. Other subcommittees focusing on statewide projects and evaluation coordination are discussed in the next two sections.

---

[5] The Technical Manual, updated in October 2010, can be found here: http://www.dps.state.ny.us/TechManualNYRevised10-15-10.pdf

Lesson Learned-Better Evaluation at a Lower Cost: A Statewide Approach

Another major activity for Staff and the EAG is to identify research projects that could be undertaken most cost effectively on a statewide or regional basis (e.g., best practices, baseline studies). It may not make financial sense, for example, for each individual program administrators to perform their own study regarding the best evaluation or program design strategies. Not only would the total cost likely be higher, but also the final products collectively may not be as strong as the utilities and NYSERDA pooling their resources to pursue a single study with a coordinated approach. The EAG created a subcommittee to develop a priority list of the most viable studies to be conducted on a statewide or regional basis. Effective and comprehensive evaluation of the EEPS portfolio should not be limited to analyses focused only on the program-specific process and impact evaluations, because this approach only captures part of the story. It is also important to examine the broader impacts of the EEPS portfolio, such as assessing market dynamics (e.g., how is the market changing?), understanding the effect of emerging technologies (e.g., use of LED lighting), and monitoring product baselines (e.g., the percentage of homes in New York with high efficiency furnaces). This type of research can help support program strategy, design and implementation, and also better document program impacts.

In May 2011, the EAG kicked off a project to develop best practices for process evaluation within New York's program portfolio. This effort is intended to provide evaluators with recommendations to facilitate common approaches and increase compatibility of results among program administrators. In the near future, the EAG expects to approve Requests for Proposals for two baseline studies designed to provide the EAG and other policy makers with a better understanding of the market saturation and the energy savings potential of many of the residential, commercial and industrial measures currently being installed through EEPS programs. This information will also help to determine the degree to which the EEPS programs have influenced the marketplace. Despite the sometimes complicated issues related to developing a project that meets the needs of the various program administrators, the efforts of the EAG's Statewide Studies Subcommittee have, so far, been driven by a spirit of cooperation and congeniality. In fact, while designing the baseline study, the subcommittee members quickly agreed on the research areas that needed to be addressed, and even the details of the scope of work fell into place fairly easily. The major stumbling block for both projects has proved to be administrative details, such as the cost share of the studies, the logistics of project/contract management, and the contract terms. Each program administrator has its own internal structure and policies and some program administrators require an individual contract with the evaluation firm. Simply stated some program administrators could not simply send a check for their share of project costs to the program administrator designated to manage the contract. Designating Staff as the contract manager was not an option because of the complexities of the state procurement and contracting process. An additional complication was the size variance of the potential contributions, when project cost allocation is based on the percent of EEPS funding allocated to each program administrator. NYSERDA, for example, represents around 65 percent of the EEPS/SBC funding and the remaining 35 percent is divided among the utilities.[6] Similarly, Con Edison has about 16 times and NYSERDA about 27 times the funding for electric programs compared to Rochester Gas and Electric (RG&E), the utility serving the Rochester area. The EAG is actively working to develop solutions to these problems.

Lesson Learned-Coordination Can Help

Another important EAG activity is to explore ways of better coordinating evaluation activities to avoid duplication of effort and increasing the overall effectiveness of the evaluation process. For

[6] NYSERDA's share currently includes an annual allocation of $175 million in System Benefits Charge (SBC) funds.

example, if several program administrators need to interview appliance distributors serving New York about stocking patterns, could the program administrators coordinate an approach to conduct the interview once and share the results? The coordinated approach would save money and time and would probably make appliance distributors more likely to cooperate. These types of efforts are under active consideration and, recently, most of the program administrators agreed to conduct a joint impact evaluation of the multiple gas furnace rebate programs operating in New York. A scope of work has been drafted and the logistical details are under discussion. A coordinated approach has the potential to result in a more accurate sample, more reliable results and lower costs. Of course, coordination may be a bigger issue in New York than in other states because of State's large number of energy efficiency programs and program administrators. However, the concept of better coordination has wide applicability.

Lesson Learned-Keep the Advisory Group on Message

In teaching Federalism, some political science instructors explain the relationships of the various layers of government (i.e., federal, state and local) as being more like a marble cake than a layer cake. This comparison is intended to highlight the numerous and complex interactions that often occur between the various branches and levels of government. The same comparison can be used to describe the relationship between energy program implementers and evaluators. While some tasks are clearly associated with one area or the other, some tasks are murky. A good example is reporting. The EAG and Staff played the lead role in developing the reporting requirements and overseeing the reporting system for the EEPS program portfolio. Staff and the EAG worked diligently to determine and define the reporting metrics and to create a process capable of serving the needs of about 100 programs being administered by 11 program administrators.

This time-consuming effort resulted in the EEPS monthly, quarterly and annual progress reports as well as the data collection protocols for evaluation purposes. A prime objective of the Staff and EAG reporting effort was to ensure that program administrators reported key performance indicators that would serve as an "early warning" to Staff when programs are deviating from their goals. Another key objective was to establish a list of data elements that are commonly needed for program evaluation and oversight activities, such as the exact type and date of measure installation. This project-tracking information is not regularly reported, but program administrators are required to collect and track these measure-specific data for all EEPS program projects.

On the surface, it might seem easy to track basic program/project milestones. However, there are many factors which can make it difficult to accurately and consistently define, collect, enter and report project and program information. For example, to track the status of an application approval, how do you deal with a program with a multi-step application process (e.g., approval for audit, approval for measure installations)? This was one of hundreds of questions that the process generated, and is still generating, about two years after the release of formal detailed reporting guidelines. [7]

Upon reflection, the development of the reporting process would have likely been more effective if there was more involvement of both Staff specializing in implementation issues and the program administrators. Their involvement would have added to the depth of knowledge of the program details. To support this point, there was some confusion over responsibility among Staff. The Staff associated with implementation issues believed that the monthly, quarterly and annual progress reports were a product of the evaluation team, and therefore it was the responsibility of the Evaluation Staff to review the progress reports, and then inform the Implementation Staff if there were any areas that required attention. Evaluation Staff, however, saw their responsibly as collecting the data, performing a

[7] The reporting guidelines can be found here: http://www.dps.state.ny.us/Reporting_Manual_6-30-09.pdf
An update is expected in the summer of 2011.

preliminary data quality check and turning the results over to the implementation staff for a more in-depth analysis and follow-up with the program administrators.

In December 2010, the Commission took action to address this type of problem by creating an Implementation Advisory Group (IAG). The Order (Case 07-M0548, 12/21/2010) stated:

> *"Many parties expressed a desire for more collaboration regarding the implementation of EEPS programs. In order to further the discussion of issues which ultimately affect the success of our EEPS programs, it is appropriate that we create an advisory group, much like the Evaluation Advisory Group (EAG) which we created in our June 23, 2008 Order in this proceeding. The EAG has proven valuable in coordinating the discussion of evaluation issues, and we would expect the advisory group we are now creating to serve a similar role with regard to program implementation issues. Therefore, the Director of the Office of Energy Efficiency and Environment is directed to establish an Implementation Advisory Group, hosted by Staff and consisting of representatives of all EEPS program administrators to advise Staff on implementation issues and to assist in program coordination among program administrators."*

Staff's reporting protocols are currently being reviewed by **both** the EAG and the IAG. Staff expects a revised reporting format to be available this summer. A longer range objective is to create a centralized reporting database system to help Staff monitor program progress and make key data, including evaluation reports, readily available to policy makers and the public via the internet. Staff expects this system to be in place in early 2012.

Staff Review of Evaluation Plans -- Lessons Learned

Staff plays a major role in reviewing, not only the evaluation plans, but also the key evaluation deliverables, such as survey instruments, sampling plans, logic models and draft reports. This is a time-consuming process, but one that Staff feels is essential to ensuring rigorous evaluations. In fact, Staff's reviews have proven especially valuable in the early stages of the EEPS effort. After all, most New York utilities were returning to administering large energy efficiency program portfolios after an absence of about 10 years, and NYSERDA needed to rethink its evaluation approach because of the significant increase in evaluation funding which allows for the use of more rigorous, but more expensive evaluation techniques. Staff needed to have a comfort level that each evaluation was being conducted in accordance with the Evaluation Guidelines. In addition, maintaining organizational separation between the program evaluation and program implementation functions mandated increased Staff oversight. Another important advantage of the Staff review function was the early identification and correction of problems in the evaluation process. Often, if a problem is uncovered late in the process, it is impossible, or at least impractical, to go back and make a correction. For example, if an important question is omitted from a survey, it is generally not cost effective to administer the survey again. Additionally, flaws in the sample design could result in biased and inaccurate results. While the quality of the evaluation deliverables has generally been good, the review process has led to recommendations that have enhanced the evaluations, such as:

- Expanding and refocusing researchable issues for a program targeting small business
- Adding actionable recommendations for program improvement as a key deliverable
- Enhancing the sampling strategy for several programs

As one utility evaluation contractor pointed out, while most other states do not examine the components of an evaluation with the same level of detail as New York, New York's reviews and related recommendations have notably enhanced the evaluation effort. It is always challenging to

balance the need for oversight and the risk of micro-managing. It is Staff's intention that, once the program administrators become more experienced with Staff expectations and Staff confidence in the evaluation products increase, the level of review will be reduced accordingly.

Program Evaluation Results -- Lessons Learned

The exact timetable for the completion of evaluation results is dependent on a number of factors, but it is often strongly driven by the number of program participants and the time required for installing specific measures. For example, for a program targeting new commercial sector construction, it is not unusual for two or more years to elapse from the initial program application to full project completion because of the complexity of the new construction process (e.g., design issues, obtaining building permits, construction delays because of weather.). For other programs, such as those targeting residential furnaces/boilers, installations are not especially time consuming and peak customer activity levels cluster around the heating season.

Ultimately, to produce meaningful results, evaluators must have sufficient data to analyze. For some types of impact evaluations, evaluators seek energy consumption data for a period of about 12 months before the installation of an energy efficiency measure, and then again for about 12 months after the measure is installed. While process evaluation can be conducted much earlier in the program cycle, it often makes sense from a statistical standpoint to wait until the population of completed projects is large enough to select a sample that is truly representative of a cross section of participants. Through May 1, 2011, Staff received final process evaluation reports for fourteen EEPS programs, draft market characterization and assessment reports for two EEPS-related programs, and draft impact evaluation reports for two EEPS programs. Staff had initially expected more evaluation results to share, but the delays associated with program implementation impacted the flow of evaluation results. As of year-end 2010, 25 programs of the approximately 100 EEPS programs were not even reporting energy savings. In addition, the utilities needed time to establish their internal evaluation departments and hire independent evaluation contractors. For example, Con Edison, the state's largest utility, has focused to date almost exclusively on process evaluation, but expects to begin impact evaluation this summer.

An important lesson learned is that evaluators need to improve communication with stakeholders regarding the availability of evaluation results. Some stakeholders assumed that evaluation would be conducted on nearly a real time basis, not fully appreciating the length of time required to conduct a comprehensive evaluation and what could be expected and when. Arguably, Staff and the EAG could have done a better job in this area. One of the major items in the EAG's strategic plan for 2011 is to tackle the issues of how best to communicate evaluation results. This will be a prominent issue in the months ahead as more evaluation reports become available. By fall 2011, Staff expects to receive process evaluations covering an additional 30-40 programs and impact evaluations for 10-15 programs (including NYSERDA's SBC programs). An important step in the effort to improve communication, the author presented a status report evaluation progress at the May 19, 2011 Commission session. The presentation covered a number of topics including a summary of initial findings from the evaluations received to date, a snapshot of EAG achievements, and the outlook for the future (Saxonis, 2011).

Staff's objective is to have the evaluation data serve as a tool in a process of continual program improvement and not as an academic report that collects dust on a bookcase. Consistent with this objective, the early evaluations have been insightful. For process evaluations, a key requirement is to identify lessons learned and provide specific actionable results for program improvement. In addition, Staff requires program administrators to include these recommendations in their EEPS quarterly reports accompanied by their response to the recommendations. The evaluation results to date indicate that many elements of the EEPS programs are working well. On the other hand, some evaluations have uncovered issues that need to be addressed, such as improving application and rebate processing, better

targeting of potential customers, and increasing training of program staff to improve responses to customer inquiries and concerns. A vivid example of the continual program improvement concept is Central Hudson's Small Commercial Program, a program designed to provide audits to non-residential customers with an electric demand of 100 kilowatts or less. This customer segment includes small businesses, local governments, not-for-profits, private institutions, public and private schools, colleges, and healthcare facilities. A tepid response to the program prompted Central Hudson to conduct its process evaluation earlier than usual. Not surprisingly, the evaluation concluded that the program was not meeting its objectives, and the evaluation offered numerous recommendations for improvement dealing with program management, database tracking, and marketing. As a result, Central Hudson "fully restructured" the program, and it is now showing significantly improved results.

Conclusion

The bottom line is that without credible evaluation the future of energy efficiency programs will be jeopardized. Energy efficiency programs must prove themselves every day, or else they will rapidly lose support not just in New York, but nationwide. It is clear that evaluation is more than a collection of formulas, statistics, and models; it contains a heavy dose of human dynamics, judgment, interpretation, and communication. It also involves the careful execution of skills related to planning, policy, and collaboration. Even the most elegant algorithm will disappoint if program administrators fail to plan for basic tasks such as collecting the required data. The effectiveness of good planning would certainly be undermined without "buy in" from policy makers. The technical dynamics and the human dynamics must synchronize. The simple lesson is to plan, anticipate needs, communicate, and effectively execute shared goals and objectives. Staff's evaluation team and the EAG are critical components in the success of New York's evaluation efforts. Staff has made solid progress over the last three years, but continues to face challenges. One immediate challenge is to carefully isolate and report energy savings to be able to determine the savings attributable to specific programs in order to better understand and monitor program progress. When consumers are confronted with multiple energy efficiency messages and programs sponsored by federal, state, and local governments, utilities, and retailers (e.g., Wal-Mart's environmental sustainability program), the quantification of success or failure of programs can be difficult to assign. Over time and based on all the combined influences that customers experience, it will be difficult to determine whether any change in energy behavior is the result of a federal government program, a CFL discount coupon from the local supermarket, or the lasting impact of a program from five years ago. Some in the evaluation community have argued that too much attention and funding is placed on determining attribution. They argue that it is too difficult, maybe even impossible, to accurately measure, and wonder whether the exact cause of the energy saving action really matters. While there is some merit to this argument, from a regulatory perspective, there is a need to know if ratepayer money is being spent wisely. It is difficult to explain to a Commissioner that we do not know if the dollars being collected from ratepayers to fund these programs during these difficult economic times are responsible for the drop in energy consumption. Staff is working on finding an answer to this difficult question.

While there is much more work that needs to be done, Staff is pleased with the results. The achievements over the last few years have helped build a strong foundation for the years ahead. The EAG/Staff Strategic Plan for 2011 includes efforts to improve reporting, refine data collection guidelines and improve the communication of evaluation results. It is all about "continual improvement." Representatives from several states, Brazil and Canada have contacted us to learn more about our evaluation efforts. As one person said, "In evaluation, we just follow what New York is doing."

References

Barakat & Chamberlin. 1993. "*Standard Reporting Format for Annual DSM Program Evaluation Results*," submitted to the Program Evaluation Task Force, New York Power Pool, Albany NY.

Consortium for Energy Efficiency (CEE). 2010. "*2010 Annual Energy Efficiency Industry Report.*" http://www.cee1.org/files/2010%20State%20of%20the%20Efficiency%20Program%20Industry.pdf

New York State Public Service Commission. June 23, 2008. Case 07-M-0548, "*Order Establishing Energy Efficiency Portfolio Standard and Approving Programs."* Albany NY.

New York State Public Service Law, Section 65, Article 4, Subsection 7.

New York State Public Service Commission. December 21, 2010. Case 07-M-0548, "*Proceeding on Motion of the Commission Regarding an Energy Efficiency Portfolio Standard.*" Albany NY.

New York State Public Service Commission. May 19, 2009. Case 07-M-0548, "*Proceeding on Motion of the Commission Regarding an Energy Efficiency Portfolio Standard*." Albany NY.

Saxonis, W. 2009. "*A New Evaluation Paradigm for a New Generation of Energy Efficiency Programs."* Proceedings of the 2009 International Energy Program Evaluation Conference, Portland Oregon.

Saxonis W. 2011. Presentation on evaluation at the May 19, 2011 New York State Public Service Commission Session. http://nydps.onlinevideoservice.com/Archive/Event_051911.aspx , See 36 minute mark.

Vine E, N. Hall, K. Keating, M. Kushler, R. Prahl. 2010. *"Emerging Issues in the Evaluation of Energy Efficiency Programs: The US Experience."* Proceedings of the 2010 International Energy Program Evaluation Conference, Paris France.