How Organizations Implement Evaluation Results

Jane S. Peters, Research Into Action, Inc., Portland, OR
Sharon A. Baggett, S.A. Baggett, Portland, OR
Patricia Gonzales, New York State Energy Research Development Authority, Albany, NY
Paul DeCotis, New York State Energy Research Development Authority, Albany, NY
Ben Bronfman, Quantec, LLC, Portland, OR

Abstract

The New York State Energy Research and Development Authority (NYSERDA) operates the **New York Energy $mart**[SM] Program which is funded through a Systems Benefits Charge (SBC). In 2001, funding for evaluation increased from .05% to 2.0% of the overall program budget and the state policymakers concurrently raised their expectations for the evaluation of the program. In 2003, NYSERDA began a significant evaluation effort using four specialized contractors to provide evaluation services.

In 2005, the Process Evaluation specialty contractor was asked to conduct an assessment of the evaluation effort. The Evaluation Review, which is the subject of this paper, examined the first three annual cycles (2003-2006) through multiple interviews with 32 contacts (including program staff, NYSERDA Energy Analysis staff responsible for the evaluations, the four specialty evaluation contractors, and the co-chairs of the Advisory Group, which is comprised of stakeholders and oversees the evaluation of the SBC-funded programs) and a literature review. The interviews focused on NYSERDA staff experience with the evaluation process over the three-years, as well as the response of program staff to the recommendations provided in the evaluations. The literature review sought to identify best practices in evaluation utilization and evaluation capacity building for comparison to NYSERDA.

The results of the Evaluation Review demonstrate that NYSERDA is building evaluation capacity within program and Energy Analysis staff, that the evaluation approach is capable of meeting program and stakeholder needs, and that evaluation utilization is increasing over time. Relative to best practices, NYSERDA has demonstrated steady progress along a continuum that is common for other large organizations.

Introduction

Beginning in 2001, funding for the evaluation of the **New York Energy $mart**[SM] Program, operated by the New York State Energy Research and Development Authority (NYSERDA), increased from 0.05% to 2.0% of the overall program budget. With this increase in funding, expectations of state policymakers involved with the State's System Benefits Charge (SBC) program and other stakeholders increased. NYSERDA's Energy Analysis program, which is responsible for conducting and overseeing evaluation efforts, added staff and took this opportunity to develop a new approach, using a portfolio structure with multi-program evaluations, mixed methods, and a different way of using evaluation contractors than is typical in traditional energy program evaluations.

In 2003, NYSERDA embarked on a comprehensive evaluation effort using four specialized contractor teams to provide services. This approach was unique among energy efficiency evaluation efforts in that each of the specialty contractor teams had a target focus: Program Theory and Logic; Market Characterization, Assessment and Causality; Measurement and Verification; and Process Evaluation. In addition, a fifth team focused on Evaluation oversight and support.

NYSERDA's evaluation efforts depart from the traditional approach of focusing on a single program. In a traditional energy program evaluation, single programs are evaluated, using any or several of the primary types of evaluation – process, measurement and verification, benefit/cost, etc. – and are completed by either a contracted evaluator or a contracted evaluator using subcontractors. Thus, the evaluator brings into the effort all of the skills needed to conduct the approaches solicited or needed for a single program evaluation, and the evaluation team becomes familiar with all aspects of the program. Development of methods (usually a mix of quantitative and qualitative), analysis of results, and synthesis of data are the responsibility of the evaluator/evaluation team. At the end of an evaluation cycle, the program provider usually combines the results from each program evaluation and "rolls them up" to the portfolio level to provide an estimate of the overall effects of the portfolio of programs.

The primary objective of the NYSERDA approach was to meet the needs of the multiple stakeholders and policymakers, and provide estimates of savings at the portfolio level. While single program evaluations were conducted, management's goal was to foster shared information, plan for collaboration, and identify and exploit synergies across programs. The expectation was to synthesize results across the different evaluation approaches to assess program effectiveness at the portfolio level, while providing some evaluation results at all other levels – sector, market, program, project, and measure – to help improve program implementation. NYSERDA's Energy Analysis group envisioned that this approach would provide advantages in terms of consistency of methods, integration of results, and coordination (Mahone et al. 2004).

In 2005, the Process Evaluation team was asked to assess the evaluation process. The Evaluation Review (Review) examined the first three annual cycles (2003-2006) of NYSERDA's evaluation efforts. The purpose of the Review was to assess the outcomes of the evaluation approach and to learn how the evaluation results were being used. The intended outcomes of the evaluation approach were to build evaluation capacity, to integrate evaluation into the program process, and to meet stakeholder requirements.

This paper discusses the results of the review of how NYSERDA used the evaluation results during the first three years of program evaluation and places that in context of other organizations' approach to evaluation. Following a discussion of the methodology, three key components of evaluation utilization – organizational learning, direct utilization of evaluation recommendations, and evaluation capacity building – provide the context for discussion of NYSERDA's evaluation experience. The authors' conclusions and recommendations complete the paper.

Review Methodology

The methodology employed for the Review, grounded in process evaluation methods and approach, was conducted in two phases – the first in 2005 and the second in 2006. In-person and group interviews, as well as a review of secondary data, were included. Table 1 shows the distribution of interviews. Across the two years, 27 of the contacts were interviewed in both phases. A total of 32 individual interviews were conducted with NYSERDA's Energy Analysis and program staff members, as well as with members of the SBC Advisory Group that oversees the evaluation effort;[1] four group interviews were held with the specialty evaluation contractor teams.

[1] The SBC Advisory Group is assigned by the New York Public Service Commission (PSC) to be the independent SBC program evaluator. The PSC, in Opinion No. 98-3, issued on January 30, 1998, specified initial SBC funding levels and duration (July 1, 1998, to June 30, 2001), and named NYSERDA as the PSC's third-party independent SBC administrator. (http://www3.dps.state.ny.us/pscweb/WebFileRoom.nsf/ArticlesByCategory/86EBE0283819224285256DF100755FE5/$File/doc3640.pdf?OpenElement). SBC operating arrangements were finalized among the PSC, New York State Department of Public Service (DPS) staff, and NYSERDA in the March 1998 SBC Memorandum of Understanding (MOU). The 1998 MOU also directed NYSERDA to solicit public input in developing its draft SBC Operating Plan for the initial three-year SBC period, and to establish an outside SBC Advisory Group.

Table 1. Interview Samples for the Review of the Evaluation Process

Data Collection Method	Number of Completed Interviews	
	2005	2006
Individual interviews with NYSERDA program staff members	20	18
Individual interviews with NYSERDA Energy Analysis staff members	9	9
Individual interviews with NYSERDA senior management	1	—
Group interviews with Evaluation contractor teams (specialty teams plus general assistance contractor)	—	4 groups
Individual interviews with representatives of the Advisory Group	—	2
Total	30 individuals	29 individuals; 4 groups

Figure 1 shows the evaluation structure from 2003-2006. The Energy Analysis group at NYSERDA is charged by the Public Service Commission (PSC) for managing the specialty evaluation contractors and coordinating efforts between program staff and these contractors to complete the evaluation activities for each program. The evaluation results are reported to the SBC Advisory Group which then transmits them to the PSC as the Independent Program Evaluator.

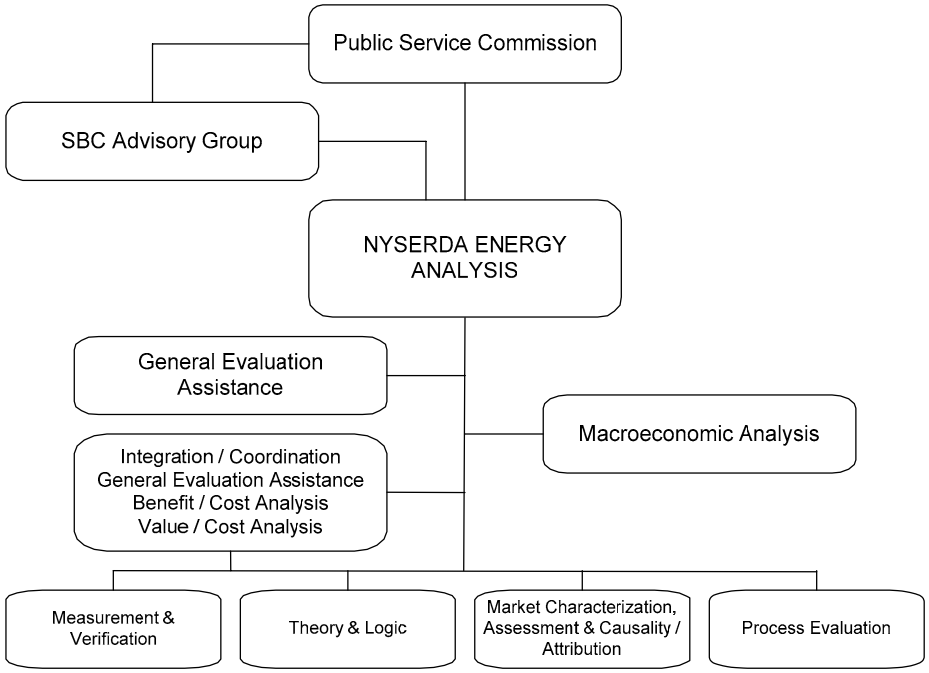

Figure 1. New York Energy $mart[SM] Evaluation Structure & Evaluation Team

Responses to the evaluation recommendations were examined during the interviews and through follow-up surveys. The Review interviews asked each contact to reflect on the three evaluation cycles and the types of evaluation efforts conducted in each. NYSERDA program staff were asked to assess the degree to which they had taken action in response to the recommendations of evaluators and to identify the reasons for their actions or inaction. A total of 244 recommendations were reviewed: 174 in 2005 for the first two evaluation cycles (2003-2005) and 70 in 2006 for the third cycle (2005-2006).

These data are qualitative; therefore, qualitative analysis software was used to code interviews and extract themes and underlying issues. To facilitate the analysis, a common set of questions was used to

address key issues: roles and responsibilities of the interviewee and their perspective on the overall evaluation process. Two steps were used in reviewing responses to the recommendations. First, the Process team identified those recommendations that were "actionable"; only these would be reviewed with program staff during the interview process. Second, after the interviews were completed, three raters coded staff responses to the questions on their action or lack of action regarding each recommendation. This was necessary because staff members frequently provided commentary on their approach to the recommendations that was inconsistent with their scoring of the recommendation as acted on or not. This process generated a program-specific matrix that illustrated whether the recommendations were addressed or not addressed, barriers to adoption of the recommendations, and whether any changes made as a result of the recommendations were considered to have improved the programs.

To place the results of the Review in context with wider practices in regard to the use of evaluation in large organizations, the Process team conducted a review of secondary literature related to evaluation utilization practices in a wide range of organizations, from energy utilities to international development agencies. The goal was to provide an overview of how other energy efficiency entities, as well as large organizations in different fields, use evaluation findings in planning, program design, and program implementation. The team also conducted a literature review of both energy conference papers and of the broader literature on evaluation and organizational policy. The best practices research sought to address the following questions:

- What is reported in energy conference proceedings and other literature on the use of evaluation recommendations?
- What are the practices for how organizations respond to evaluation recommendations?
- Do organizations have any "standards" for how evaluators should frame or structure recommendations?
- To what extent do policies on recommendations, accountability, and other variables affect the degree of evaluation utilization and resulting organization and program change?

The Context of Evaluation in Organizations

Evaluation fits within organizational learning. The concept of organizational learning means building upon what works and exhibiting a willingness to change, probably continuously, in order to be more effective. For organizational learning to occur, key components are needed: very clear goals, priorities, and measures; and an alignment of goals and measures for people (staff), budgeting and financial management, and for assessing organizational results. Also, very importantly, there must be clear communication among people working in the organization, as well as with stakeholders, partners, and customers.

Most organizations, especially public agencies, usually have accountability as a clear and stated goal. Organizational learning may not be clearly identified or included in the concept of accountability. However, these two are compatible because in learning, organization accountability becomes one tool for continuous improvement. Without clear outcomes and measures for each, staff is left to determine where to place the emphasis, and confusion regarding task importance is evident. Evaluation is thus one part of the overall strategic process – where members are involved in collaborative, communicative, and reflective evaluation processes, organizational learning will occur (Preskill & Torres 2000).

Keys to successful organizational learning are the conditions supporting it. Those identified consistently throughout the literature include: systematic support, empowering (including allowing risk-taking), communication (sharing), collaboration, and shared vision. Under these conditions, evaluation should be learning-centered, offering diverse chances to share and develop evaluation participants' knowledge, skills, and attitudes (Song 2002).

There are also conditions under which evaluation information can better enhance organizational learning. Two conditions that are of particular value for this assessment are *evaluation utilization* and *evaluation capacity building*.

How Organizations Respond to Recommendations – Evaluation Utilization

The term *evaluation utilization*, or *evaluation use*, refers to the degree to which the findings of evaluation influence the programs, processes, or policies investigated, or affect the organization in which these occur. Evaluation use is of interest because public organizations explicitly or implicitly hold the view that systematic inquiry about organizational activities should influence decision-making. The degree to which evaluation information is disseminated and used varies widely, and there is a great deal of discussion within the evaluation profession on both conditions to improve utilization, as well as the kinds of utilization that are legitimate (and valuable) outcomes of evaluation (Dahler-Larsen 1998; Forss, Rebien & Carlsson 2002; Grasso 2003; Hodges & Hernandez 1999; Mark 2004; Valovirta 2002; Weiss 1998). Several types of evaluation use are found (Owen 1999; Weiss & Bucuavalas 1980) including:

- *Instrumental use* – evaluation directly influences decision-making or the program itself
- *Conceptual use* – evaluation influences thinking about the program, even if it does not lead to decision making
- *Legitimative use* – evaluation is used to justify decisions made and respond to concerns of policy-makers for continued support
- *Process use* – cognitive and behavioral changes result from the user's involvement in the evaluation process (enhanced skills in evaluation, routine incorporation of evaluation in the organization's processes, and program improvement – also termed capacity building)

In this study, we primarily addressed instrumental, conceptual, and process use, assessing the extent to which evaluation results were used to directly influence program processes or thinking about the program, as well as the degree to which NYSERDA's Energy Analysis and program staffs' capacities in evaluation were enhanced.

Recommendations for programmatic or organizational changes resulting from evaluations fall within instrumental and conceptual use, i.e., where evaluation directly influences decision-making or the program itself by changing implementation or the way the program is conceptualized. Measuring the consideration and use of evaluation recommendations is one of the most straightforward ways to assess whether the organization's staff values the evaluation results and whether the results are providing useful direction for program improvement. The literature discusses the types and levels of recommendations made by evaluators, whether they should be made at all, and, if so, in what format, and how the organization evaluated the extent to which recommendations are made from evaluation findings. Yet within this, the team uncovered little in the way of standards for use of evaluation recommendations, and thus relied on an informal poll conducted in 2006, which suggested a wide range of practices in energy evaluation.

The informal poll was conducted by TecMarket Works, which sent an email to the 25 members of the planning committee of the International Energy Program Evaluation Conference (IEPEC 2006) asking for assistance in identifying whether any organizations have a policy requiring that program administrators or managers must respond to evaluation findings and report whether or not they are implementing the recommendations. A lively discussion ensued regarding the presence of such requirements and their value. Responses of practices regarding evaluation recommendations ranged from a formal policy to no policy. Examples along this continuum include:

- *Bonneville Power Administration (Oregon)*: Evaluation teams are required to make formal recommendations. The implementation teams are required to respond in writing to the Vice President, including response, action plan, and a timeline for action, if necessary.
- *Wisconsin Department of Administration (Focus on Energy)*: Evaluations require specific recommendations; program administrators are expected to respond. This was termed a quasi-formal process, where face-to-face discussions are held to review recommendations.
- *California Public Utilities Commission*: Evaluation teams are required to make recommendations; this is considered a primary deliverable. No response is required from the program staff.
- *Energy Trust of Oregon, Inc.*: Program staff are not required to make changes in response to recommendations, but are required to address them in a response memo, which is a public document appended to the evaluation report. Program staff reviews of final reports are also limited to input/changes on matters of fact, not to opinions or recommendations of evaluators.
- *National Grid (Northeast U.S.)*: No requirements.

Despite this range of responses, the majority of organizations reported that recommendations were expected from evaluations, but there were few or limited requirements for response to the recommendations on the part of program staff and administrators.

NYSERDA's Response to the Evaluation Recommendations

The evaluation specialty contractors were asked to include recommendations at the conclusion of each study. The resulting recommendations proved to include actionable recommendations, as well as recommendations that were more comments on current practice. Excluding those on which actions could not be taken resulted in 244 recommendations over the three evaluation cycles. Several examples of recommendations illustrate the range among the different specialty evaluation contractors:

- Staff should continue to streamline the application process by allowing incentive application requirements to be fulfilled electronically, and staff should explore the possibility of spot-checking selected requirements for "proven" installers rather than checking every application (MCAC).
- As part of the ongoing database quality assurance and control activities, NYSERDA should review all pre-qualified measure records for consistency between the units of the stipulated savings and the parameters tracked for each measure (i.e., count, size, or capacity), and correct any errors that are found. Tracking records were moved to the Buildings Portal database in June 2006; these errors should be corrected before they become perpetuated in the new platform (M&V).
- Develop a set of multifamily case studies based on projects conducted in NYC that describe existing project benefits and lessons learned from the perspective of owners, building managers, and tenants (Process).

Table 3, at the end of this paper, displays the responses to questions asked about what actions had been taken on these 244 recommendations. As can be seen in the table, staff reported taking some action on 48% of the recommendations in the first two cycles and on 67% of the recommendations in the third cycle. The differences in the two periods of time reflect some of the changes that have been occurring within the evaluation process. A possible source of the increase is that the evaluation specialty contractors have gained

increased knowledge and understanding of the **New York Energy $mart[SM]** programs operated by NYSERDA. The evaluation specialty contractors noted improved communication with program staff, which affected the evaluation questions to be addressed and thus the focus of any changes or recommendations arising from the evaluations. Program staffs, however, generally did not report a perception of improved recommendations on the part of the evaluation specialty contractors. Program staff generally do not think there is any requirement for them to respond to the evaluation recommendations, although some do state they should and others clearly state that they want to use the evaluation results. While the SBC Advisory Group specifically had requested information on how program staff responds to evaluation recommendations in 2004, program line staff seemed generally unaware of this expectation.

These findings for NYSERDA suggest that in comparison to the results of the informal poll, NYSERDA falls somewhere near the mid-range of other energy organizations in how it addresses evaluation recommendations. As NYSERDA continues through this type of review process to explicitly track responses to recommendations, then it will be moving towards greater evaluation instrumental utilization than most other energy organizations. Currently, recommendations are required of evaluation contractors, as is most common, and there is an implicit expectation, but not a formal policy that staff will respond to these recommendations. Responses may include good reasons for inaction (e.g., program changes since the evaluation make the recommendation no longer viable, lack of funding or staff capacity to implement, or it is not efficient to implement until a new program cycle).

How Organizations Implement Evaluation – Evaluation Capacity Building

Evaluation *capacity building* according to Stockdill et al. (2002) refers to "the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its uses routine." Evaluation capacity building is seen as a way to use evaluation findings to: assist in planning, decision-making, and prioritization (especially in the budgeting process); assist managers by revealing program performance that will lead to future learning and improvement; assist with accountability; and demonstrate the extent to which activities have been successful (Mackay 1999).

A recent review of best practices by Baizerman et al. (2002) on evaluation capacity in large organizations analyzed the evaluation efforts at the Centers for Disease Control (CDC), the American Cancer Society, the World Bank, and a large school district. In the more successful organizations, such as the American Cancer Society and the World Bank, critical elements include a vision, clear goals, plans, and what Baizerman et al. describe as an environment that is "purposive, attentive, and reflective of work with others in the organization in order to create and sustain ways of ensuring that evaluation studies and their uses continue to be carried out (105)."

Successful evaluation-capacity building organizations have a clear guide, incorporate evaluation in their everyday work, have enough funding to support ongoing evaluation studies, and create the conditions that support the request and use of evaluation findings. Thus, there are site structures, cultures, and practices that enhance evaluation capacity building. (These are similar to those noted in the above review of practices essential for utilization.)

Among the site structures needed to ensure evaluation capacity building are an identifiable process, structure, and resources. The process should have explicit public and executive support, the strategy should be identifiable, evaluation staff should be at the table when decisions are made, and the evaluations should be demand-responsive. Baizerman et al. (2002) note that the needed cultural factors are a common positive understanding that evaluation is "how things are done," a culture of inclusiveness of many disciplines and all stakeholders, a transparent and participatory evaluation practice, and ongoing efforts to create a language and conversation regarding how evaluation works. This should be combined with making evaluation capacity building explicit – *explained and promoted actively and appropriately throughout the site.* This

would also include ensuring that the evaluations conducted are of high quality and in accordance with commonly accepted standards.

A recent study by Taut (2007, 56) confirmed much of this with a findings that "for widespread use of self-evaluation and evaluation for learning to occur: 1) staff must own the evaluative process and have responsibility for the quality of the evaluated project, and have the autonomy to bring about suggested changes; 2) the organizational culture as well as the immediate work environment must be characterized by trust, transparency, and a constructive approach to mistakes and failures; 3) endorsement and modeling of learning from evaluation by organizational leaders and senior-level managers is necessary to set the tone that learning from evaluation is valued; and 4) sufficient resources and time must be dedicated to the evaluation practice."

How NYSERDA Staff Responded to the Evaluation Process

Understanding of Evaluation Purpose, Roles and Requirements

Evaluation capacity building is a process. NYSERDA had a vision, had set forth goals, and had a basic plan. While almost the entire NYSERDA program staff members interviewed in 2006 thought they were sufficiently informed of the purpose and requirements of the evaluation activities, when asked specifically to rank in order of importance of two possible purposes of the evaluation effort – to provide information to stakeholders and to provide information to improve programs – views diverged, demonstrating that the communication of the goals and plans has been incomplete. As one program staff person noted during the interview, "Where is the theory and logic model of the evaluation?"

Other areas of importance in building an effective evaluation approach are clear roles for the actors (staffs and evaluators), as well as well-defined expectations for providing necessary data and other information for reporting. For many of those interviewed, both of these were slow to develop. Among program staffs, Directors, and the Energy Analysis staff, almost all agreed that there are no formal requirements for how much time and effort program staff should devote to an evaluation.

Working with Specialty Evaluation Contractors

One of the key sources of dissatisfaction between program staff and the specialty evaluation contractors was the time it took for the contractors to fully understand the programs and having to repeat the program information to so many different evaluation teams. Other challenges focused on disagreement about the measures proposed to reflect program outcomes (especially in the Research and Development programs) and what some program staffs perceived as a lack of creativity among the contractors, who were seen as relying solely on "the same old utility-based approaches" that do not adequately address market transformation efforts, a focus of many of the programs.

Lack of coordination among the evaluation teams, although improved over time, was also a challenge for program staffs. The process evaluation team asked program staffs and members of the Energy Analysis staff to rate the coordination among the specialty evaluation contractors in the first year and in the third year of the contracts. Their rankings are shown in Table 2. As seen, the ratings improved over the course of the evaluation contracts, with many more of the staff members of both types rating the specialty evaluation contractors as "very coordinated" or "somewhat coordinated" in Year 3. In addition, one Director noted that early on, it seemed that the specialty evaluation contractor teams were not used to working together (more often being competitors for projects), did not understand or feel comfortable in NYSERDA's more collaborative approach to report development (many drafts, input from multiple sources), and were not preparing the evaluation information in a way that decision-makers could receive it.

Table 2. Ranking of Coordination Among Specialty Evaluation Contractors

Ranking of Specialty Evaluation Contractors	By Energy Analysis Staff		By Program Staff		Total	
	Year 1	Year 3	Year 1	Year 3	Year 1	Year 3
Very coordinated	1	2	0	4	1	6
Somewhat coordinated	1	3	3	8	4	11
Somewhat uncoordinated	4	1	6	3	10	4
Very uncoordinated	0	0	5	0	5	0
Total Number of Staff Responding[1]	6	6	14	15	20	21

[1] Numbers responding in Years 1 and 3 are not equal, as some staff said they were not involved enough in Year 1 to provide a rating.

Increased success in coordination among the specialty evaluation contractor teams, NYSERDA's program staffs, and the Energy Analysis staff occurred during the three years reviewed. For example, members of the specialty evaluation contractor teams cited as indicators of success the following: program staffs' adoption of the logic model process; co-authorship of conference presentations; and improvements in data-tracking systems. NYSERDA's program staffs, Directors, representatives of the SBC Advisory Group, and the Energy Analysis staff reflected that the specialty evaluation contractors brought many of the skills needed to achieve NYSERDA's goals. As one program staff person generalized, "I think they did a pretty good job. They tried to find creative and effective ways to assess the programs." Over the three years, Energy Analysis staff increased their technical and management capabilities and was increasingly able to manage the evaluation reporting process that had initially been a part of the Evaluation Oversight contract.

Building Evaluation Capacity

In building evaluation capacity, there have been many challenges, largely due to funding constraints and initial time constraints as NYSERDA's program and Energy Analysis staffs had to not only get programs up and running, but also initiate an evaluation system. The system chosen was a complex evaluation approach with lower funding than is common for energy organizations,[2] and there was inconsistent communication of the evaluation plan at all levels of the organization. The Energy Analysis staff had a more complete idea of the evaluation approaches, and the importance of these approaches for reporting, but this message was not clearly understood by most program staffs. While Energy Analysis held annual meetings to review the previous cycle, prepare for the coming year, and discuss successes and challenges, no such review or preparation was reported by program staffs.

In addition, most members of the Energy Analysis staff were new to evaluation and initially did not have the experience in evaluation theory, implementation, measurement, and analysis to be more than conduits between the specialty evaluation contractors and program staffs. In many cases, data systems were not adequate for reporting, confusion existed over reporting timelines and schedules, and the roles of all involved were not clearly defined.

Positive Change Over Time

These challenges have been and are being addressed as the evaluation of NYSERDA's programs has evolved. From the specialty evaluation contractors' perspectives, there have been positive changes.

[2] As noted above, NYSERDA allocated 2% to evaluation; many energy organizations allocate 4-6%, with 10% a historical best practice of the Bonneville Power Administration in the early to mid-1980s.

Program staffs seem to feel less threatened by the effort and to place higher value on evaluation, and the contractors' knowledge of the programs and coordination has improved. Contractors also noted that the fact that NYSERDA is conducting this evaluation review demonstrated a strong desire to continue to improve the evaluation process. The Advisory Group representatives also noted increased clarity regarding the role of the Advisory Group in evaluation, certainly improving greatly from the first funding cycle through the second, and are pleased with the evaluation results.

From the program staffs' views, the specialty evaluation contractors have begun to learn more about their programs and to assist them more, as does the Energy Analysis staff. Program staffs noted they are identifying evaluation issues at the beginning of program cycles, their data tracking systems are improving, and the importance of the evaluation effort for public reporting is better understood. The Energy Analysis staff is becoming more knowledgeable and skilled, and roles are becoming more defined and settled. Energy Analysis staff, who work with the evaluation contractors and coordinate their work with program staffs, also noted that there had been an increase in interest in evaluation on the part of program staff and that the evaluation specialty contractors were more knowledgeable and informed about the programs as the evaluation cycles progressed. While not all of the challenges have been addressed, with the exception of a few program staff members, most of those interviewed noted sometime during the interview process that positive changes have occurred.

Conclusions

The results of the Review suggest that between 2003 and 2006, NYSERDA has made significant progress in creating an effective evaluation process with limited resources. All of the groups represented in the interviews indicated that, while year one was very difficult, the second and third years were greatly improved in terms of communication and overall coordination between program staffs, the Energy Analysis staff, and the specialty evaluation contractors. Significant improvements were also made or begun on program databases to better support the evaluation effort.

The Review indicates that evaluation capacity has increased among program and Energy Analysis staffs. While some of the specialty evaluation contractors and others believe there is still room for increased knowledge of evaluation theory and practice among the Energy Analysis staff, many comments in this review focused on the improvement of team members in the more practical aspects of evaluation – surveys, interviews, etc. – and measures and understanding of data requirements.

Among program staff, evidence of increased capacity, as well as evaluation utilization, is demonstrated by use of the evaluation findings and incorporation of evaluation thinking into program planning (and doing so earlier in the process) and implementation (through solicitations, data collection, guidance for implementation contractors, and involvement of specialty evaluation contractors in process changes). Even where program staff did not use the information to change programs, increased awareness of processes, justification of existing practice, or delaying action now but holding the possibility of future action, were all mentioned as outcomes. These changes show a significant growth in understanding of the value of evaluation among program staff.

Lack of guidance regarding the recommendations has been present throughout the period evaluated and continues. In spite of a direct request from the SBC Advisory Group for reporting on action taken or not taken on each of the recommendations, the emphasis placed on responding to the recommendations varies by Program Manager. While an assessment of action taken on recommendations is met through the Review, greater clarification for program staffs about the expectations for the recommendations is needed. NYSERDA is at the exemplary end of the continuum by requiring evaluators provide recommendations, but lies somewhere around average compared to other energy organizations' practices regarding response to recommendations (e.g., desired but no mandatory requirement for written response).

The evaluation effort was most successful in accomplishing the third, and perhaps most important, outcome desired – meeting stakeholders' expectations – with the primary stakeholders: the SBC Advisory Group, the Public Service Commission, the Department of Public Service, and the executive and legislative branches of the New York State government. The SBC Advisory Group reported high levels of satisfaction with the evaluation effort, indicating that program evaluation, as well as the programs themselves, exceeded their expectations. Stakeholder satisfaction is also demonstrated by the approval of increased funding for NYSERDA, as the SBC administrator, for a third cycle.

Program staffs are another, albeit less critical (in policy perspectives) stakeholder, and for them, the evaluation was perceived as less helpful in meeting their expectations. While program staff reported improvements over time in the quality and relevance of the evaluation findings, there are still strongly held views that the results are not sufficiently timely, do not reflect the full impact of program implementation, reflect the use of incorrectly applied measures, are based on samples too small to draw conclusions, or are not actionable. NYSERDA is taking action to address some of these issues.

The results of the Review point to opportunities for improving the evaluation process at NYSERDA and demonstrate the value in conducting such a study. The recommendations of the Review were presented to the SBC Advisory Group, as well as to the Energy Analysis and program staffs. NYSERDA has learned that they have a process that is working and evolving in a positive direction and recognizes the opportunities to improve the process. The recommendations of the Review suggested that NYSERDA develop a revised an expanded evaluation plan, which they have now done. This plan should be clearly communicated to program staff and include a formal statement of how to respond to evaluation recommendations. NYSERDA has begun that process this year. Additionally, NYSERDA plans to continue to conduct evaluation review activities.

References

Baizerman, M., D.W. Compton, and S.H. Stockdill. 2002. "Summary and Analysis of the Case Studies: Themes Across the Cases." *New Directions for Evaluation* 93 (Spring): 101-107.

Dahler-Larsen, P. 1998. "Beyond Non-Utilization of Evaluations: An Institutional Perspective. *Knowledge, Technology and Policy* 11 (1&2, Spring/Summer): 64-90.

Forss, K., C.C.Rebien, and J. Carlsson. 2002. "Process Use of Evaluations." *Evaluation* 8 (1): 29-45.

Grasso, P.G. 2003. "What Makes an Evaluation Useful? Reflections from Experience in Large Organizations." *American Journal of Evaluation* 24 (4): 503-514.

Hodges, S.P. and M. Hernandez. 1999. "How Organizational Culture Influences Outcome Information Utilization." *Evaluation and Program Planning* 22: 183-197.

IEPEC. 2006. Email correspondence with members of the IEPEC Planning Committee, of which three of the authors are members: Ben Bronfman, Paul DeCotis and Jane Peters.

Mahone, D., P.A. DeCotis, H. Kim, and K. Chappell. 2004. "A Revolutionary Approach to Evaluation: NYSERDA's Portfolio Model." *In Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings*. Washington, D.C.: American Council for an Energy-Efficient Economy.

Mackay, K. 1999. "A Diagnostic Guide and Action Framework." *Evaluation Capacity Development* (ECD Working Paper Series No.6). Washington, D.C.: World Bank.

Mark, M. 2004. "The Mechanisms and Outcomes of Evaluation Influence." *Evaluation* 10(1): 35-57.

Owen, J.M. 1999. *Evaluation: Forms and Process*. Newbury Park, Calif.: Sage.

Preskill, H. and R.T. Torres. 2000. "The Learning Dimension of Evaluation Use." *New Directions for Evaluation* 88: 25-37.

Song, Y. 2002. "Creation and Utilization of Evaluation Information for Organizational Learning." *In Academy of Human Resource Development Conference Proceedings*. Baton Rouge, La.: Academy of Human Resource Development.

Stockdill, S.H., M. Baizerman and D.W. Compton. 2002. "Toward a Definition of ECB Process: A Conversation with the ECB Literature." *New Directions for Evaluation* 93 (Spring): 7-25.

Taut, S. 2007. "Studying Self-Evaluation Capacity Building in a Large International Development Organization." *American Journal of Evaluation* 28(1): 45-59.

Valovirta, V. 2002. "Evaluation Utilization as Argumentation." *Evaluation* 8(1): 60-80.

Weiss, C. 1998. "Have We Learned Anything About the Use Of Evaluation?" *American Journal of Evaluation* 19(1): 21-34.

Weiss, C.H. and M.J. Bucuavalas. 1980. "Truth Tests and Utility Tests: Decision Makers Frames of Reference for Social Science Research." *American Sociological Review* 45 (April): 302-313.

Table 3. Reported Program Staff Response to Recommendations: Combined Three Year Results

Outcome	Residential & Low-Income				Business & Industrial				Research & Development				Total by Cycles			
	2003-2004 & 2004-2005		2005-2006		2003-2004 & 2004-2005		2005-2006		2003-2004 & 2004-2005		2005-2006		2003-2004 & 2004-2005		2005-2006	
	Freq	%	Freq[2]	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
Action taken	24	44%	17	68%	35	37%	21	51%	5	21%	2	50%	64	37%	40	56%
Action taken; felt already doing it	—	—	—	—	6	6%	3	7%	4	17%	—	—	10	6%	3	6%
Partial action	1	2%	1	4%	6	6%	3	7%	1	4%	—	—	8	5%	4	6%
Subtotal of actions taken	25	45%	18	72%	47	49%	27	65%	10	42%	2	50%	82	48%	47	67%
Considered; felt already doing it	9	16%	2	8%	24	25%	1	2%	5	21%	—	—	38	22%	3	2%
Considered; no action	10	18%	5	20%	7	7%	11	27%	6	25%	2	50%	23	13%	18	28%
Not considered	11	20%	—	—	17	18%	2	5%	3	13%	—	—	31	18%	2	4%
Subtotal of actions not taken	30	54%	7	28%	48	50%	14	34%	14	59%	2	50%	92	53%	23	33%
Total[1]	55	99%	25	100%	95	99%	41	99%	24	101%	4	100%	174	101%	70	100%

[1] Total not equal to 100% due to rounding.

[2] Responses to two recommendations were "don't know"; so a total of 27 responses recommendations were received.