

Evaluation of Public Benefits Programs: A Review of the NYSERDA Experience to Date

*Paul DeCotis and Patricia Gonzales, New York State Research and Development Authority
Jane S. Peters, Research Into Action, Inc.
Sharon Baggett, Sharon A. Baggett Associates
Luisa Freeman, PA Consulting*

ABSTRACT

The New York Energy Research and Development Authority (NYSERDA) administers the **New York Energy SmartSM** public benefits programs for the State of New York. In 2005, at the close of the second year of formal monitoring and evaluation reporting, NYSERDA requested that its process evaluation contractor conduct a systematic review of the evaluation process itself to understand the degree to which program staff have acted upon the recommendations from evaluators, their reasons for action/inaction, their view of the evaluation process itself and its usefulness to them for attaining program improvements. Also important to management was to assess the extent to which the evaluation function is integrated into program design and implementation, and how the evaluation process is perceived by NYSERDA program staffs and stakeholders. The process evaluation contractor conducted interviews with 30 NYSERDA program and evaluation staff members and identified recommendations for improving the evaluation process.

Introduction

The **New York Energy SmartSM** programs are funded by an electric distribution System Benefits Charge (SBC) paid by customers of Central Hudson Gas and Electric Corporation, Consolidated Edison Company of New York, Inc., New York State Electric and Gas Corporation, National Grid, Orange and Rockland Utilities, and Rochester Gas and Electric Corporation. Programs are available to all electric distribution customers who pay into the SBC. The New York State Energy Research and Development Authority (NYSERDA), a public benefit corporation established in 1975, has administered the SBC funds since 1998. As a result of its successful programs and NYSERDA's ability to report program progress toward its public benefits goals, SBC funding in New York has been reauthorized twice, each time for an additional five years with an increase in funding. In addition, in light of NYSERDA's successes, it now administers the statewide renewable portfolio standard program, as well as two utility natural gas efficiency programs and a targeted system-wide, demand-side management program for Consolidated Edison.

NYSERDA Evaluation Structure

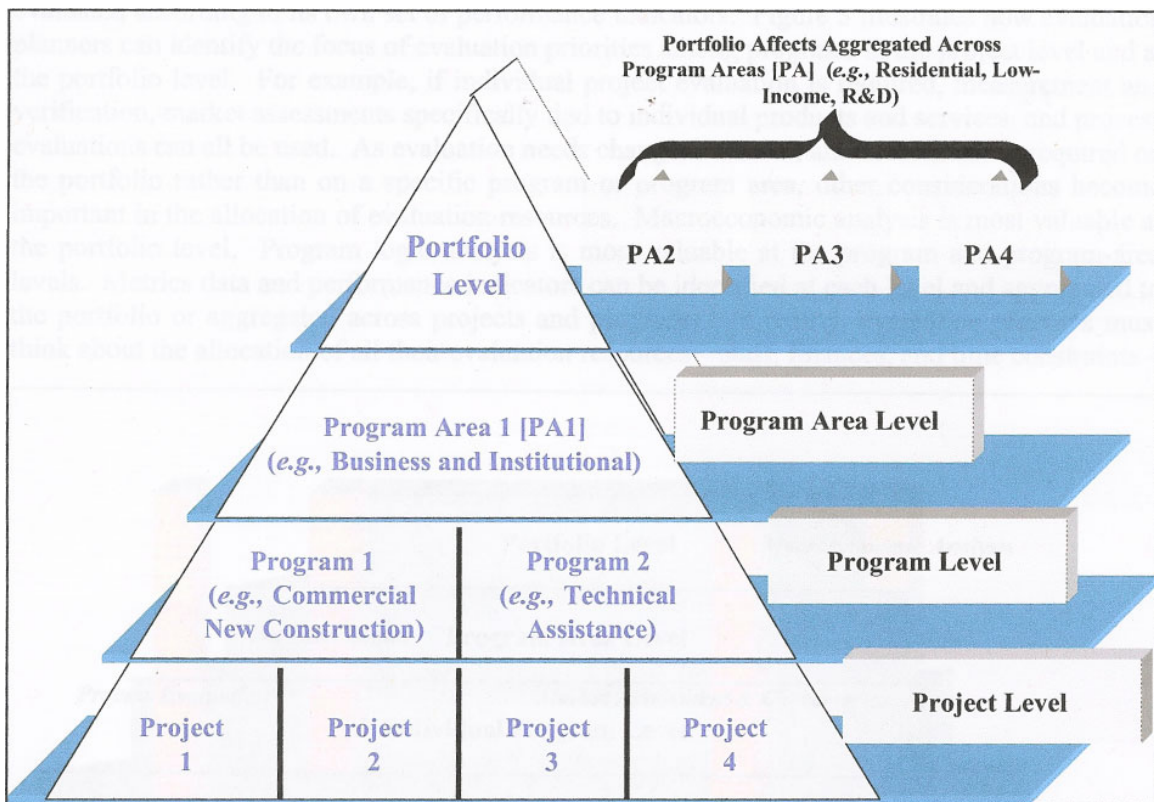
There have been two rounds of SBC funding to date, one from 1998-2002 and a second from 2003-2006. The third round of funding was approved in 2006 and extends through 2011. In the second round of funding, NYSERDA's evaluation budget increased from 0.05% to 2.0% of the overall program support. Expectations of state policy makers and other stakeholders

increased with the expansion of funding. Staff at NYSERDA took this opportunity to develop a portfolio approach to evaluation that involves a different way of using evaluation contractors than is typical in energy program assessment. The overall goal is to combine multiple program evaluations in an effort to determine the extent to which the overall public policy goals of the **New York Energy SmartSM** Program are being met.

The change in the organization of the evaluations was to shift from comprehensive assessments of each program, conducted by a single team knowledgeable about it, to specialized teams that would provide evaluation services to the entire range of programs offered by NYSERDA. The four specializations were to include measurement and verification evaluation (M&V), market characterization, assessment, and causality evaluation (MCAC), process evaluation, and program theory and logic models (PT/LM). For any particular program, the M&V team would assess gross savings, the MCAC team would assess net-to-gross and market effects, the process evaluation team would assess the program processes, and the PT/LM team would develop a logic model and a program analysis to support it.

Over a period of several years, all of NYSERDA's programs would receive some level of evaluation review. Larger programs with the greatest reported energy and peak savings impacts might receive annual reviews across all evaluation specialties, smaller programs might only have an M&V or a process evaluation completed. The management goal in using this approach was to foster shared information and collaboration, as well as to exploit naturally occurring and manufactured synergies across programs. The larger evaluation goal was to provide evaluation results at the portfolio level. This aggregation to the portfolio level is shown in Figure 1.

Figure 1. NYSERDA's Portfolio Evaluation Structure



Source: DeCotis & Winters 2003

NYSERDA staff envisioned that this approach would provide several advantages (Mahone et al. 2004):

- better integration of evaluation results across the portfolio of programs;
- greater consistency of evaluation approaches;
- fewer evaluation contracts to administer;
- better coordination among evaluators; and
- direct access for program staff to knowledge and skills of evaluation experts that would influence future program planning, design, and implementation.

To implement this model, each specialty evaluation contractor began the evaluation period (2003-2006) by working with NYSERDA evaluation staff to prioritize programs and identify potential projects to assess, as well as to determine the type and level of evaluation to be conducted. The specialty contractors then brought their assessments together and through a collaborative process determined: 1) whether any evaluation should be conducted; and 2) what approach(es) would be used.

On an annual basis, some programs are not evaluated, others may be evaluated using only one or two approaches (e.g., process only, process and MCAC), and some receive the full range offered. Over the full four-year cycle, however, most programs experience at least one evaluation effort.

The primary objective of the portfolio evaluation approach is to provide policy makers with the data and information necessary to assess the progress toward goals of the **New York Energy SmartSM** Program. Thus, the reporting and presentation of results is required to be policy relevant, high-level enough to easily comprehend and engage in discussion, rigorous to the extent necessary for the results and reporting to be trusted, and balanced – with successes and disappointments equally reported.

Multi-Level Evaluations

Looking at Figure 1, it is obvious that the NYSERDA portfolio, like most energy efficiency portfolios, has multiple levels at which evaluation questions and issues can arise. There are measure-level issues such as persistence and durability, there are project-level issues such as occupant behavior and changes in occupancy levels, there are program-level issues of implementation and delivery, and there are sector-, market-, and portfolio-level issues. At every level, estimates of savings can be framed differently. For most energy organizations, the savings are “rolled up” from the project and program level to the portfolio level. In NYSERDA’s case, with a limited evaluation budget and high expectations for reporting, the objective was to provide sufficient confirmation of savings, implementation, and market effects to verify the portfolio-level results.

There are many approaches to and reasons for using multilevel evaluations. As Yang et al. (2004) note, in the non-profit sector, national foundations use cluster evaluations to evaluate a program being administered at different, autonomous sites with the intent of bringing about a common change. In the international development arena, multilevel country program evaluations are used to assess the results of a donor’s funding in strategic areas: for example, rural development, reducing gender equalities, or improving natural resource management, all of

which are aimed at reaching the larger goal of strengthening the capacity of the poor (IFAD 2004).

In studies of social and economic programs (Banks et al. 2002; Gondolf 1998; Rog & Randolph 2002), multi-site evaluations have been used to examine a continuum of program approaches, to obtain adequate sample sizes, and to increase the ability to generalize research conclusions. In education, where many national models are adopted, or mandated over large geographic areas, or in complex systems, evaluation efforts most similar to the one described here are being evaluated using theories of change. In a theory-of-change model, high-level strategy outcomes are identified, from short- to long-term, and these are developed as part of a performance indicator framework generated prior to project start-ups (Rogers, 2005).

NYSERDA's approach, while being a different approach for energy efficiency organizations, tracks closely with the variety of multi-level approaches being taken by evaluators of other social programs. DeCotis and Mahone (2005) observed several benefits of the NYSERDA solution to the multi-level evaluation. One was a sense of ownership of the evaluation process among the evaluation contractors – contractors are thinking creatively and proactively about ways to further improve the evaluation of NYSERDA's portfolio. Contractors are also working with program staff to improve methods and databases for program tracking that will not only make their evaluation work more efficient in the future, but will also provide better management data for the programs. However, as with other multi-level evaluation approaches, there are lingering issues and concerns about how to integrate findings. It is in within this context that NYSERDA hoped to gain insights on the effectiveness and usefulness of their portfolio evaluation approach.

Approach

NYSERDA prepares an annual evaluation report that includes the findings for each evaluation activity, as well as conclusions and recommendations. As of spring 2005, when the body of research being reported in this paper was conceived, two cycles of monitoring and evaluation reporting on **New York Energy SmartSM** programs had been completed. The recommendations for improvements in the programs and the evaluation methods had been included in the two evaluation reports.

After the first cycle (2003-2004) was completed, program staff prepared a written response to each of the evaluation contractors' recommendations. For the second (2004-2005) cycle, the combined recommendations were compiled from the annual evaluation report by the process evaluation team.

Combining the various recommendations provided a total of 223. Of these, the process evaluation team selected the 174 actionable recommendations to be the focus of the interviews with staff. Thus, the first component of the evaluation review was an assessment of responses to the 174 actionable recommendations made by the five evaluation contractors over the two-year period.

In addition to asking staff about the status of the recommendations, the process evaluation team developed a series of questions to assess NYSERDA program and evaluation staff's experience with the evaluation process. The staffs were also asked to provide suggestions as to how it might be improved. Managers and officers were further asked to reflect on the usefulness of the evaluation for public reporting.

NYSERDA has over one hundred program staff, although not all are directly involved in evaluations. Thus, the process evaluation team conducted interviews with twenty program staff, nine evaluation staff, and one corporate officer.

The in-depth interviews were analyzed using qualitative data analysis procedures, including identification of themes and emergent trends. The evaluation team used qualitative analytic software, *Atlas ti*TM, which is a powerful tool for assigning codes to narrative text, for counting codes, for examining where key topics are mentioned singly or in unison with others, and for facilitating comparison of narrative themes across types of respondents. In addition, the team compiled a matrix table for each program reviewed, showing the recommendations examined, actions taken to date to address the recommendation, and influences, where identified, on actions taken or not taken by program staff.

Findings

The analysis of the interviews resulted in findings on the evaluators' 174 recommendations and the evaluation process.

Findings on Recommendations

Each recommendation was reviewed with staff members, during which time they were asked a series of questions, including:

- Was the recommendation considered?
- If considered, was it acted on?
- If acted on, did this result in program improvement?
- If not acted on, what were the barriers to doing so?

Responses to these questions permitted each recommendation to be coded regarding the degree of action taken or not taken. For those recommendations from the 2003-2004 cycle that already had a staff response, the evaluation team assigned an outcome code after reviewing the comments. The analysis resulted in six categories, three for *action taken* and three for *no action taken*.

The process evaluation team found that some action was taken on 48% of all actionable recommendations presented. This seems to be a reasonable degree of response for the first two years' of effort; recognizing that staffs had, in some cases, less than 12 months to address recommendations for programs already in the field, and that good reasons might exist for not acting on a recommendation. NYSERDA's program and evaluation staffs both noted that the evaluation process has continued to improve over the period of time covered by this research.

The findings from the analysis of these responses, by sector are shown in Table 1.

Table 1. Action on Recommendations

Outcome	Residential & Low-Income		Business & Industrial		Research & Development		Total	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
Action taken	24	44%	35	37%	5	21%	64	37%
Action taken; felt already doing it	—	—	6	6%	4	17%	10	6%
Partial action	1	2%	6	6%	1	4%	8	5%
Subtotal of actions taken	25	45%	47	49%	10	42%	82	48%
Considered; felt already doing it	9	16%	24	25%	5	21%	38	22%
Considered; no action	10	18%	7	7%	6	25%	23	13%
Not considered	11	20%	17	18%	3	13%	31	18%
Subtotal of actions not taken	30	54%	48	50%	14	59%	92	53%
Total¹	55	99%	95	99%	24	101%	174	101%

¹ Total not equal to 100% due to rounding

In reviewing Table 1, each category can be interpreted in the following manner:

- *Action taken* – clearly the recommendation was acted on and recommended changes were made to programs.
- *Action taken; felt already doing it* – a clear action was taken, but staff members commented that they felt they either had been taking the action on their own or were proceeding toward taking the action before the evaluation results and recommendations were published.
- *Partial action* – some portion or aspect of the recommendation was acted on, but not all, for any variety of reasons.
- *Considered; felt already doing it* – staff members commented that while they had considered the recommendation, they felt they were already doing the action, which the evaluators might not have been aware of or misunderstood; therefore no new action was taken.
- *Considered; no action* – the staff members commented that they had read and understood the recommendation and its implications; they had considered it, but either perceived barriers that limited their ability to act, or they did not agree with the recommendation.
- *No action taken* – staff comments indicated that they did not consider the action at all or felt that the recommendation was not directed at the program staff level, but pertained either to the evaluation process or to NYSERDA management; in either case, staffs felt the recommendation was not theirs to act upon.

Findings on the Evaluation Process

Interviewees at all levels noted that they see evidence of greater incorporation of “evaluation thinking” into program design and implementation. All recognized that the first and second year evaluations were times when both the evaluation teams and NYSERDA staff were

learning the portfolio-level evaluation approach and that by the time of this review during the third year, some of the difficult challenges had been resolved or had decreased substantially.

As noted previously for the **New York Energy SmartSM** programs in general, there are as many as five levels of stakeholders with different and potentially very important evaluation needs and issues that could be addressed through an evaluation process. These levels include: the portfolio as a whole; sector issues, such as for the residential programs and their associated market; program issues resulting from the design and implementation approach selected; project-specific issues; and specific questions about measures, such as savings and persistence.

The primary purpose of NYSERDA's evaluation efforts is to report on portfolio-level accomplishments in order to provide policy decision-makers with the data and information they need to assess the efficiency and effectiveness of the **New York Energy SmartSM** Program relative to its established public policy goals. NYSERDA also hopes the evaluation information is sufficient to improve program design and service offerings. The evaluation approach therefore included both single-program evaluations and market-level studies. This led to expectations by program staff that they would gain detailed information about their programs and markets.

Despite the apparent focus on a single program, the single-program evaluations were generally designed to produce outcomes that could be used for portfolio-level reporting, which left some program staffs disappointed that they did not receive information from the evaluations that would help them improve their programs, learn more about their markets, or gain other levels of expertise desired. Some program staffs were also left dissatisfied with the perceived constraints placed on the evaluations because the cost-effectiveness methodologies, which were developed to meet requirements of the Public Service Commission (PSC), did not capture the full impacts of their programs (e.g., market transformation, education, non-energy benefits, etc.). Even though the cost-effectiveness evaluation included several cost-effectiveness scenarios that do capture the fuller impacts from programs, there were some program staffs who did not feel the multiple scenarios fully or accurately reflected the outcomes of their programs.

Another challenge for the evaluation was uncertainty about the role and responsibilities of the NYSERDA evaluation and program staffs. The evaluation staff had a more positive perception of the value and usefulness of the evaluation for program efforts than did program staffs. Program staffs more often saw it as challenging to deal with NYSERDA evaluation staff as a liaison to the evaluation contractors.¹ These differences of perspective reflect confusion over the various actors' roles and responsibilities, as well as the purpose of the evaluation, on the part of program staff. It was clear that NYSERDA program staff, evaluation staff, and the evaluation contractors all have expectations of what the other would do in order to accomplish the evaluation. Although management reports that discussions were held regularly between evaluation and program staffs regarding the evaluations, the review found that program staffs still had many concerns about the evaluation process and its purpose.

Finally, the evaluation contractors did not get off easy. Program staff perceived much of the evaluation process as burdensome. Program staff, as well as the process evaluation team, also noted that some of the recommendations made by the contractors were not actionable, at least not by program staff, or they were imprecise as to whether the recommendation should be addressed by the program staff or by supporting departments within NYSERDA. The lack of clarity in some of the recommendations also led some program staff to be uncertain as to whether the

¹ NYSERDA's evaluation staff plan the evaluations, manage the specialty evaluation contractors, and also serve as a member of each contracting team as the "on-site" liaison between program staff and the contractors.

evaluation teams fully understood the programs being administered by NYSERDA and the New York environment in which the programs are being implemented.

Conclusions and Recommendations

The evaluation review found that the evaluation process has support within NYSERDA and that there are opportunities to improve it. Though overall SBC funding will be increased in the next period, the anticipated funding for evaluation is expected to stay at 2%. Therefore, it is clear the focus on portfolio-level evaluation needs to continue. Program staffs, however, have non-portfolio level needs for evaluation that exceed the capacity of the available funding. This review shows that program staffs have looked to the evaluation effort to help address their needs, perhaps unrealistically.

Similarly, there were significant expectations placed on the evaluation staff by the program staff. In other energy organizations, staff responsible for monitoring and evaluation activities typically have experience with evaluation, prepare the requests for evaluation services in which they define the evaluation questions to be addressed, and often define the research design for the project, and then interact closely with contractors to determine how the evaluation approaches can best benefit the program staff and other stakeholders.

In NYSERDA's case, the evaluation contractors and NYSERDA's evaluation and program staffs collectively defined the evaluation, in effect, creating the processes, data collection means, surveys, sampling methods, and tasks. Despite efforts to clarify the purpose of the evaluation and the value that the evaluation would bring to NYSERDA and to program staff, there is still work to do to ensure that all program and evaluation staff are clear on their roles and responsibilities. As a result of the process evaluation conducted that provided the basis for this paper, several recommendations emerged:

- *Recommendation:* A clear understanding of the evaluation approach and needs should be better communicated within NYSERDA and with its stakeholders, with an understanding that some evaluation activities are necessary for portfolio-level work and that some program funding might be necessary to supplement the evaluation budget in order to accomplish activities that will have greater value to program staffs.
- *Recommendation:* As part of a review of roles, NYSERDA should work to continue developing the skills of all staffs in evaluation so that communication is clearer between program and evaluation staff, and so that additional evaluation capabilities can be developed.
- *Recommendation:* Using the model of the portfolio as having multiple levels, NYSERDA evaluation staff should convey to program staff just what questions the evaluation effort will address, what questions it will not address at each level, and how the information will be used.
- *Recommendation:* Evaluators should follow four guidelines when crafting recommendations.
 1. Recommendations should clearly state what action is to be taken and who is the likely responsible party.
 2. The feasibility and relevance of recommendations should be discussed with program staff prior to finalizing the evaluation report, and then be reflected in the recommendations.

3. In designing evaluation research, be sensitive to the burden placed on customers.
4. Use of “continue to” recommendations should be limited to those cases where staff may be considering stopping actions believed to be central to program success or to those cases where emphasizing the value of current actions is very important.

Future Directions

As a result of this initial inquiry, this body of work will continue to develop. The assumed and unexpected benefits and the challenges of using this approach can be explored regularly to ensure that the evaluation is as effective as possible. The goal is to evaluate the evaluation function in a manner similar to the way programs are evaluated – to improve the efficiency and effectiveness of evaluation and its usefulness as an integrated strategy and management tool.

References

- Banks, S., McHugo, G.J., Williams, V., Drake, R.E., and M. Shinn. 2002. “A Prospective Meta-Analytic Approach in a Multisite Study of Homelessness Prevention.” *Conducting Multiple Site Evaluations in Real-World Settings: New Directions for Evaluation* 94: 45-59.
- DeCotis, Paul A., and D. Mahone. 2005. “An Integrated Portfolio Approach To Evaluating Energy Programs: New York’s Experience.” In *Proceedings of the 2005 International Energy Program Evaluation Conference*. Madison, Wisc.: International Energy Program Evaluation Conference.
- DeCotis, Paul A. and R.G. Winters. 2003. “Portfolio Evaluation: Strategy and Planning, Part I.” In *Proceedings from the 14th National Energy Services Conference and Exposition*. Phoenix, Ariz.: Association of Energy Services Professionals.
- Gondolf, E.W. 1998. “Do Batterer Programs Work?: A 15-Month Follow-Up of a Multi-Site Evaluation.” *Domestic Violence Report* 3: 64-65, 78-79.
- (IFAD) International Fund for Agricultural Development. 2004. *Arab Republic of Egypt, County Programme Evaluation*, Report No. 1658-EG. Washington, D.C.: International Fund for Agricultural Development.
- Mahone, D., DeCotis, Paul A., and H. Kim. 2004. “A Revolutionary Approach To Evaluation: NYSERDA’s Portfolio Model.” In *Proceedings of the ACEEE 2004 Summer Study on Energy Efficiency in Buildings*. Washington, D.C.: American Council for an Energy-Efficient Economy.
- Rog, D.J., and F.L. Randolph. 2002. “A Multisite Evaluation of Supported Housing: Lessons Learned From Cross-Site Collaboration.” *Conducting Multiple Site Evaluations in Real-World Settings: New Directions for Evaluation* 94: 61-72.

- Rogers, P. 2005. *Evaluation of the Stronger Families and Communities Strategy: Final Evaluation Framework*. Available online:
[http://www.facs.gov.au/internet/facsinternet.nsf/VIA/sfsc/\\$FILE/SFCS_Framework.pdf](http://www.facs.gov.au/internet/facsinternet.nsf/VIA/sfsc/$FILE/SFCS_Framework.pdf).
Melbourne, Australia: RMIT, Collaborative Institute for Research, Consulting and Learning in Evaluation.
- Yang, H., Shen, J., Honggao, C., and C. Warfield. 2004. "Multilevel Evaluation Alignment: An Explication of a Four-Step Model." *American Journal of Evaluation* 25(4): 493-507.