New Evaluation Framework: Turning the Silo on its Side[1]

Cherie Gregoire, Jennifer Meissner, Patricia Gonzales, and Victoria Engel, New York State Energy Research and Development Authority, Albany, NY

ABSTRACT

This paper describes the nimble evolution of an evaluation framework that succeeds in providing a credible and comprehensive evaluation of a portfolio of energy efficiency, demand reduction and renewable energy programs. The new model provides timely information to stakeholders, managers, and staff about the impacts, efficiency, and effectiveness of program implementation; progress in transforming constituent markets; the economic impact and cost-effectiveness of programs; and progress toward public policy goals.

Prior to 2003, evaluation was conducted by a small internal staff with limited assistance from two external contractors. With an influx of program funding, the evaluation effort was expanded to use four contractor teams to provide: program theory and logic; market characterization, assessment, and causality; measurement and verification; and process evaluation services. Rather than the typical model of one contractor performing all evaluation services segregated by program or sector, each team had a single evaluation function across all sectors. The premise was that evaluation experts in a particular field are better suited to ensure that the successes and shortcomings of diverse programs are accurately, appropriately and objectively measured and reported. This approach provides several advantages:

- better integration of results
- more consistent approaches
- fewer evaluation contracts to administer
- improved coordination between evaluators

In 2006, the number of specialty evaluation contractor modalities was reduced to three to address: impact evaluation; process assessment and evaluation management; and market characterization and assessment evaluation. This constitutes a streamlined approach while still maintaining the key building blocks of a comprehensive evaluation.

Background

First funding cycle: 1998-2001

The **New York Energy $mart**[SM] Program, New York's public benefits program, was established by order of the New York State Public Service Commission (PSC) in January 1998.[2] In this Order, the PSC designated the New York State Energy Research and Development Authority (NYSERDA) as the statewide administrator of most of the program funds and required that NYSERDA also be responsible for conducting evaluations to measure the success of the Program in meeting four public policy goals set forth by the PSC (see Figure 1).

[1] The views expressed in this paper are those of the authors and do not necessarily reflect the views of the New York State Energy Research and Development Authority.

[2] New York State Public Service Commission. In the Matter of Competitive Opportunities Regarding Electric Service, Opinion No. 98-3. *Opinion and Order Concerning System Benefits Charge Issues*. Issued and effective January 30, 1998. Cases 94-E-092 *et al*.

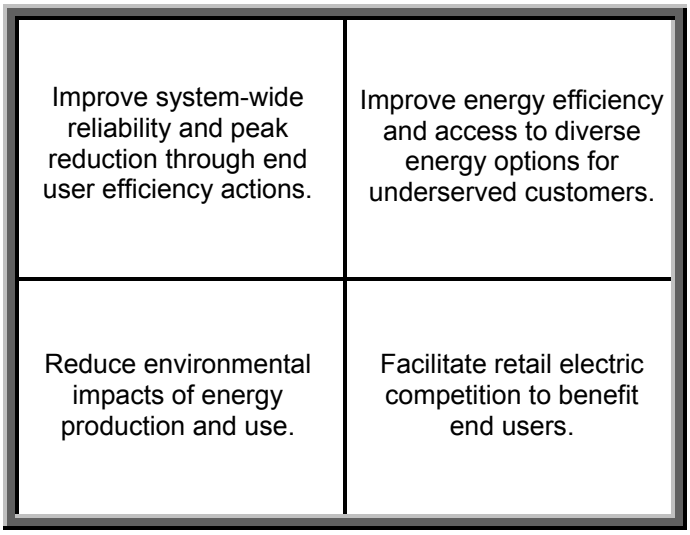

Figure 1. Public Policy Goals of the New York Energy $mart[SM] Program

The PSC provided NYSERDA with approximately $650,000 to conduct evaluation activities on the **New York Energy $mart[SM]** Program during the initial three-year program. With these limited funds, NYSERDA selected two external contractors to assist with the evaluation effort. Oak Ridge National Laboratory was selected in late 1998 to assist NYSERDA's in-house evaluation team in developing the initial evaluation framework and baseline information. GDS Associates, Inc. and its subcontractor, Megdal and Associates, were selected in early 2000 to assist NYSERDA with data collection, analysis and report writing. Other evaluation activities undertaken by NYSERDA's in-house evaluation team and these contractors were rather limited in scope and fell into three broad categories: process evaluation, outcome evaluation, and causality attribution.

Evaluations are conducted by NYSERDA on behalf of the Systems Benefit Charge Advisory Group (Advisory Group)[3], which was designated by the PSC as the independent program evaluator. As NYSERDA evaluation staff is evaluating NYSERDA implementation programs, steps have been taken to address the potential conflict of interest presented. Through the hiring of independent third-party evaluators, NYSERDA evaluation staff serves to plan, guide and manage major work efforts, rather than conducting it. Implementation staff is involved in the planning and review of results. While the contractors benefit from the internal staff knowledge of program implementation, final reports are products of the third-party evaluators. The reports are reviewed by both internal evaluation and implementation staff, yet any changes are subject to the evaluation contractor's concurrence. This compares to other organization's more one-sided approaches that limits interaction with implementation staff.

Second Funding Cycle: 2001 - 2006

In January 2001, the PSC extended the public benefits program for an additional five years, through June 30, 2006.[4] At the same time, evaluation funding was also increased to approximately $15 million over the five-year period. As per the PSC's Order and a Memorandum of Understanding,

[3] The Advisory Group consists of 24 individuals representing varied interests, including utilities, business and environmental groups, energy services companies, community organizations, professional and trade associations, and national energy efficiency and research and development (R&D) organizations.

[4] *Order Continuing and Expanding the System Benefits Charge for Public Benefits Program*. January 2001. New York State Public Service Commission.

NYSERDA was to continue evaluating the **New York Energy $mart[SM]** Program on behalf of the Advisory Group, with the Advisory Group continuing to serve as the independent program evaluator.

Even with the increased evaluation funding, the evaluation staff concluded that a portfolio-level approach, rather than program-by-program evaluations, would best serve stakeholder needs. Therefore, a major decision faced by the evaluation staff was whether to divide responsibilities by sector (residential, commercial/industrial, low-income, R&D) or by evaluation function (process, impact, logic, markets). The advantages and disadvantages associated with these approaches are included in Figure 2. NYSERDA issued competitive solicitations in fall 2002 to hire contractors to assist in evaluating the Program. NYSERDA's competitive process mandates that contractors are selected by a panel of internal and external technical experts who rate and score proposals submitted to NYSERDA.

	Model A Multiple Contractors Responsible for One Sector		Model B Multiple Contractors Responsible for One Evaluation Function
Advantages	• Each evaluation contractor receives in-depth knowledge of one sector (residential, low-income, commercial/industrial, R&D). • With greater knowledge, evaluation contractors have better ability to conduct evaluations that address several programs. • NYSERDA evaluation staff responsible for a sector becomes familiar with different types of evaluation (*e.g.,* process, impact).		• Evaluation contractors and NYSERDA's in-house evaluation staff gain knowledge of NYSERDA's myriad programs. • NYSERDA's in-house evaluation staff becomes expert in one evaluation function.
Disadvantages	• Evaluator's experience is limited to one NYSERDA sector. • Opportunities to coordinate evaluation activities across sectors are limited.		• NYSERDA's in-house evaluation staff has less opportunity to learn about other evaluation functions. • Multiple contractors would need to learn about NYSERDA's programs and there could be overlap.

Figure 2. Evaluation Models Considered

NYSERDA adopted an evaluation model consisting of multiple contractors where each contractor is responsible for a single function (Model B). The specialties required to meet NYSERDA's functional needs fell into four categories: measurement and verification; market characterization, assessment and causality; program theory and logic; and process evaluation. In addition, a general assistance contractor was procured to help with coordinating the teams and writing reports. This model was expected to provide the most value to the key stakeholders, NYSERDA staff and the evaluation contractors.

Over the past five years, this model has served NYSERDA well. Program implementation staff has become more engaged in the evaluation process over time and have gained confidence in results. In-house evaluation staff and the evaluation contractors have gained a better understanding of NYSERDA's programs, and in-house staff, in particular, has become intimately familiar with the evaluation function they manage.

Planning and Design of New Framework

In anticipation of possible renewed funding, the evaluation team seized an opportunity to take a fresh look at the framework. Planning for the new framework began in December 2005, 6 months before the end of the funding period, and the planning consisted of several brainstorming sessions held with the NYSERDA evaluation staff, manager and director. The same month, the PSC approved the continuation of the public benefit programs for 5 years at a total of $875 million, 2% of which funds evaluation. This relatively small percentage dictated that the framework be further streamlined, while

still assuring NYSERDA met PSC requirements. In the Order, the PSC specifically referenced the successes of programs illustrated in annual evaluation reports as justification for program continuation.[5]

Before considering what changes were possible, the evaluation team discussed the elements of the design that had worked well. One of the benefits of NYSERDA's framework is consistency of methodology. For example, with a single causality / attribution evaluation contractor, freeridership and spillover are measured consistently (where appropriate) across programs as well as years, allowing NYSERDA to compare these factors and to roll up progress to the portfolio level. This also allows NYSERDA to test out different methodologies based on lessons learned from earlier evaluations. The framework also allows integration of findings from the various contractors that have evaluated the same program in order to validate and strengthen individual results and present a more complete picture of program progress. An integration meeting is held annually where specialty evaluation contractors share their results and findings with the other teams. In theory, this approach would allow NYSERDA to leverage the expertise and different perspectives of the team members to provide a fuller evaluation. In practice, however, integration has been difficult to achieve with the reporting schedule. Also, of a more practical matter, NYSERDA administers 25 energy efficiency and renewable energy programs funded under the **New York Energy $mart**[SM] Program and the full-time evaluation staff numbers only seven. Designing the framework by specialty area rather than program or sector has allowed the evaluation team to manage fewer contracts and to gain an in-depth understanding of the area they manage. Given the benefits provided, going into the latest round of funding, the NYSERDA evaluation team decided to maintain the framework by specialties, rather than program or sector.

In designing the framework, the evaluation team first weighed important attributes of an evaluation, in order of priority, the attributes were:

1. Fair, objective, and credible
2. Focused on priority metrics, impacts and outcomes
3. Clearly defined purpose of evaluation activities and requirements
4. Able to demonstrate progress towards goals at both the program and portfolio levels
5. a. Meeting the various needs of stakeholders
   b. Transparent process and reporting (succinct and easily understood)
6. Clearly defined roles and responsibilities
7. Efficient administration for evaluation team and program staff
8. Thorough, comprehensive, complete
9. Integrated (inter & intra program)

This prioritization allowed the team to assess how well the current framework was meeting these attributes and how modifications could address any shortcomings. The framework designed under the last funding cycle possessed many of these attributes and had shown improvement over time.

The functional needs of the evaluation, as determined by NYSERDA's internal and external stakeholders, covered the broad areas of impact evaluation, market intelligence, customer satisfaction, cost-effectiveness, and program-level design and implementation. In addition, the team identified necessary support functions of data management, tactical general assistance, and report preparation assistance. These functional needs had been fulfilled by the four specialty contractors and two general assistance contractors.

Under the existing framework, the team recognized that key, related areas had been split between contractors. This had forced contractor teams that are normally competing for work in the market, to coordinate their work and schedules closely in order to meet reporting deadlines, while juggling their own priorities. Methodologies had to be monitored so that the results of one evaluation would be useful

[5] Case 05-M-0900, In the Matter of the Systems Benefit Charge III, Order Continuing the Systems Benefit Charge (SBC) and the SBC-Funded Public Benefit Programs, issued and effective December 21, 2005.

as the input for another team. Also, the teams make efforts to coordinate data collection in order to maximize site visits and surveys and to minimize survey fatigue. These factors led the NYSERDA evaluation team to rearrange tasks into three areas of evaluation 1) impact assessment, 2) market characterization, and 3) process assessment and evaluation management services plus a dedicated survey house. This new framework is illustrated in Figure 3. The solicitations for these services were released in August 2006 and can be found on NYSERDA's website.[6]

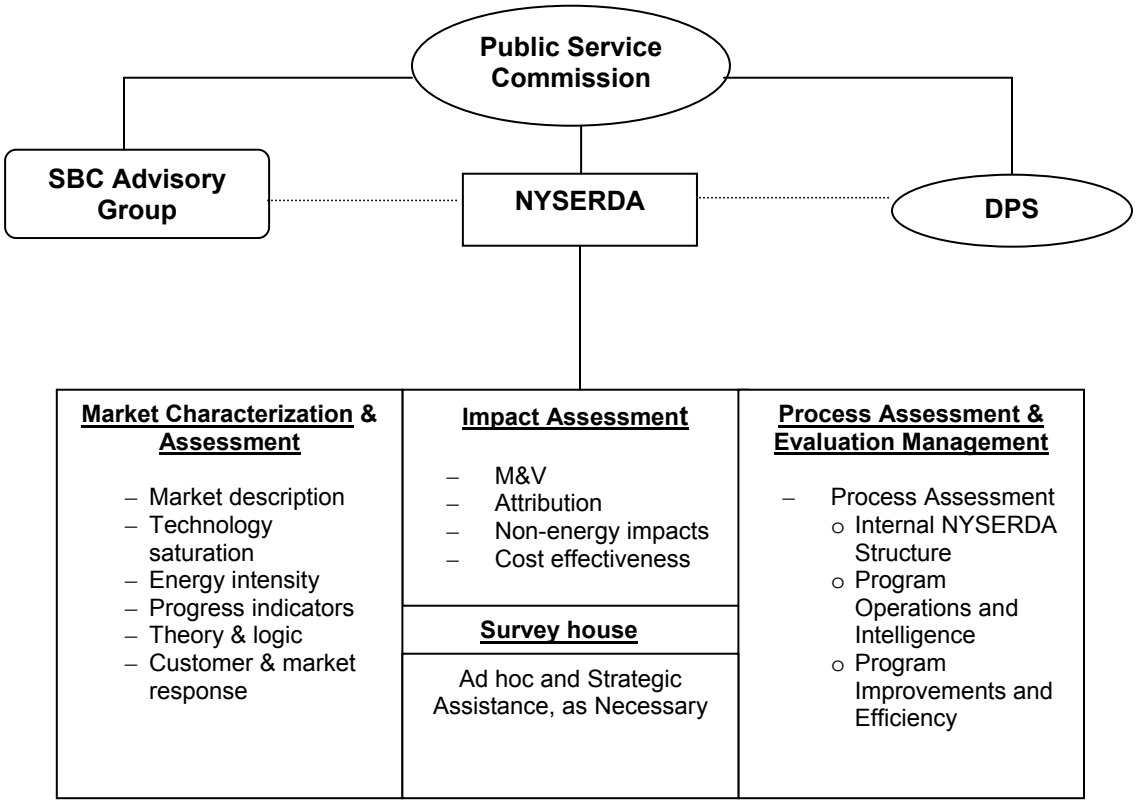

Figure 3. Evaluation Framework

Within the existing framework, savings reported by the programs are verified by the Measurement and Verification team, adjusted for any free-ridership or spillover that is identified, then entered into a cost-effectiveness tool. Recognizing the close timing and interplay of each of these steps, the framework pulls these tasks together under a single impact evaluation team that will also assess non-energy impacts. These functions had formerly been carried out by three separate contractor teams.

Formerly, program theory and logic models and progress indicator lists were developed under one team and subsequent research was performed by the Market Characterization, Assessment and Causality (MCAC) team. This created a discontinuity in the process, weakened ownership in the end product, and prevented the team from drawing firm conclusions on the position of the program in the market relative to the logic. Logic models will now be developed and the resultant indicators measured under the new Market Characterization and Assessment contract, creating a more seamless end product.

The Process Assessment and Evaluation Management team will be addressing the program process and support functions described above. Specifically, the team will be tasked with evaluating program oversight and operations; provide recommendations for program and process improvements;

[6] These closed solicitations can be found on NYSERDA's website, http://www.nyserda.org/Funding/funding.asp?i=4, under the headings for RFP 1061, RFP 1063, RFP 1065 and RFP 1086.

and serve as evaluation liaison and advisor to the NYSERDA evaluation staff and contractors. These responsibilities expand upon the tasks formerly conducted by the Process Evaluation team and the general assistance contractor. Given the experience gained by the internal evaluation team and the streamlined structure of the new evaluation framework, the need for general assistance has diminished.

In the past, individual specialty teams had their own survey firms as subcontractors. Under this new framework, most of the survey fielding functions were extracted from individual contracts and placed with a single survey house. While the lead for the specific program evaluation will still be responsible for crafting the survey, most survey fielding will be conducted by the survey house. Also, the survey house hired has experience in survey development and sampling procedures that can be leveraged. This method will potentially allow NYSERDA to realize economies of scale, while still receiving a quality product. Also, having a single survey house will provide NYSERDA with a more consistent survey administration protocol and data outputs.

Benefits of Evaluation Activities

Program planning

This section highlights several areas where the evaluation findings have contributed to program planning.

In addition to reporting on findings from program evaluations, the specialty contractors are also asked to make recommendations to improve program operations. A review of the evaluation process[7] revealed that program staff are increasingly accepting recommendations and incorporating them into implementation, reflecting the increased familiarity that the evaluation specialty contractors have with programs.

Under the current funding cycle, several programs are being consolidated and new programs are being designed. This situation provided an opportunity to develop logic models to assist in program design before the program launched.

The involvement by the same contractor across programs has afforded NYSERDA insight into possible corporate-level changes. The Measurement and Verification team has the benefit of analyzing program databases across the portfolio of **New York Energy $mart^{SM}** programs. This has revealed problems including missing data, incorrect use of stipulated savings data, etc. NYSERDA has addressed these issues by developing a single online database that houses commercial energy efficiency program data and another database dedicated to residential program data. This reflects a corporate-wide commitment to improve data storage and retrieval to improve program implementation and assist evaluation.

Sometimes, evaluation results have positive unintended consequences. For example, a residential energy efficiency program operated by NYSERDA had relied on potential energy savings to market their program to participants. Under work conducted by the MCAC team on the non-energy impacts of this program, such as comfort, they found that these factors were just as important as the energy savings in the homeowner's decision to upgrade the efficiency of their home.[8] As a result, the program modified the marketing message to emphasize the non-energy impacts.

[7] Research into Action, Inc., "Review of the Evaluation Process", NYSERDA, February 2007.

[8] Summit Blue, "ENERGY STAR® Labeled Homes and Home Performance with ENERGY STAR® Phase 1 Market Characterization, Assessment, and Causality (MCAC) Evaluation", NYSERDA, July 2004.

Organizational Strategic Planning

In addition to the traditional roles for evaluation outlined above, members of the NYSERDA evaluation team are routinely involved in corporate strategic planning. Annually, NYSERDA develops a Strategic Plan[9] describing the current strategies being employed to reach goals, as well as challenges and opportunities being pursued. Given the depth and breadth of exposure to program implementation that the evaluation requires, staff is in a unique position to take a broad view of NYSERDA goals and strategies, and a perspective on possible synergies. More directly, NYSERDA uses evaluation findings to demonstrate the efficacy of strategies at achieving progress towards goals. Also, in developing the Operating Plan[10] that allocated budgets and provided a detailed description of programs planned for the current funding cycle, past program achievements were used in establishing goals for the new funding cycle.

Lessons learned from the evaluation

In parallel with our discussion to modify the evaluation framework, NYSERDA tasked Research Into Action (the **New York Energy $mart**[SM] Process Evaluation Contractor) with conducting an evaluation review[11] covering the following key questions concerning the evaluation framework:

Was the evaluation process effectively created? The results of the evaluation review suggest that NYSERDA has made significant progress over the past few years in creating and implementing an effective evaluation process. However, an overarching evaluation plan could further improve the evaluation design and implementation. An overall vision for the evaluation was articulated as part of the March 2006 Operating Plan, and specific plans for each major evaluation project were developed with the pertinent evaluation contractors. Development of a well-defined, overarching plan for the evaluation process and outcomes could help to reduce stakeholder uncertainty about evaluation expectations and clarify the balance between evaluation aimed at program improvement and evaluation for accountability to NYSERDA's stakeholder. NYSERDA is currently developing an overarching evaluation plan, core components of which will be:

- Specific goals and measures for the evaluation
- Clear roles and expectations for the actors involved
- Clarification of the audiences to be served by the reporting process, which reporting types best serve each, and which can be produced, given budgets and timelines
- Guidelines and support for implementation of the findings and recommendations for various audiences

Once the overall evaluation plan has been developed, NYSERDA will work to clearly communicate the plan at all levels of the organization. The evaluation review suggested that clear communication of the plan could greatly reduce any uncertainty and discomfort felt by those involved in the effort. This communication will address expectations of the internal evaluation team, program staff,

[9] NYSERDA (2006), "Facing Energy Challenges in the 21st Century, A Three-year Strategic Outlook", June 2006.
[10] NYSERDA (2006). Systems Benefit Charge. *Proposed Plan for New York Energy $mart Programs* (2006 – 2011), as amended, March 2, 2006.
[11] Research into Action, Inc., "Review of the Evaluation Process", NYSERDA, February 2007.

and contractors in the model, and expectations for how recommendations from the evaluation are to be used.

Did the evaluation process have the outcomes intended? Some of the intended outcomes of the evaluation process include building evaluation capacity, greater integration of evaluation into program processes, and meeting stakeholder requirements.

The evaluation review indicates that evaluation capacity was increased to the greatest extent among NYSERDA's evaluation staff but also to some degree among NYSERDA program staff. Among program staff, evidence of increased capacity, as well as integration of evaluation into program processes, is seen in their increasing use of the evaluation findings into program planning and implementation.

Improved guidance regarding how program staff should address evaluation findings and recommendations could significantly improve the integration of evaluation into program processes. As noted earlier, this is being addressed as part of the overarching evaluation plan. Additionally, evaluation contractors now better understand the programs and are, therefore, able to: provide recommendations with more credibility and relevance; identify more actionable areas for improvement; and deliver the recommendations in a context and language that increases the receptivity of program staff.

The evaluation effort has been very successful in meeting most stakeholders' expectations. Up to this point, the SBC Advisory Group and the executive and legislative branches of the New York State government have been the primary stakeholders. While the Advisory Group has cited areas for improvement, they have consistently reported high levels of satisfaction with the evaluation effort and acknowledged that given the evaluation budget, the resulting work more than met the stated requirements. Stakeholder satisfaction is clearly demonstrated by the approval of NYSERDA as the administrator for a third cycle with increased funding. However, program staff is another key stakeholder group that would like improved timeliness and relevance from evaluation efforts. NYSERDA is taking action to address these issues. The main challenge is meeting disparate stakeholder needs, in priority order, within the available budget for the evaluation (which is smaller than in many other jurisdictions).

Is the evaluation model effective? The chosen evaluation model must be viewed in the context of whether it is workable within the requirements of the PSC and the level of funding. The key stakeholders are quite satisfied with the evaluation results. Therefore, the changes made to date and additional modifications that are under consideration focus largely on making the process less stressful and more useful for program purposes, increasing its value to NYSERDA as a learning organization, increasing evaluation capacity, and increasing its value in communicating the results to the citizens of New York.

As NYSERDA develops its overarching evaluation plan, this activity will include a more focused process in determining which programs will be evaluated and the types of evaluations to be undertaken. Among the criteria to be used in this process are program budgets, number of participating actors and customers served, program risk, stakeholder interests, and time lapsed since last evaluation. While still crosscutting, the new streamlined framework will not automatically integrate evaluation across specialties on a program basis. That is, it does not reflect an examination of each program, whether it will receive impact, process, or market evaluation(s), what issues will be examined for each, and what issues will be integrated at what levels (for sector and portfolio results). This is another overarching task which NYSERDA will undertake in this next evaluation cycle.

The evaluation framework has proven to work well for NYSERDA and has provided consistent results. NYSERDA's ability to award multi-year contracts has been crucial in the evaluation of a large number of programs with a relatively small budget. This has allowed contractors time to become

familiar with programs, for program staff to get comfortable with the contractors evaluating their programs, and for evaluation staff to learn alongside the contractors without the burden of rebidding contracts. Given the time required to become familiar with program operations, without the ability to award multi-year contracts, NYSERDA would have been limited to performing program by program evaluations, as opposed to the more comprehensive portfolio-level evaluation being conducted. In addition, this approach has given the evaluation team the opportunity to increase their institutional knowledge.

Conclusion

NYSERDA endeavors to continually improve its evaluation design, implementation, and utilization, keeping it on the forefront of energy program implementation and evaluation. The original framework design, and recent refinements, allows NYSERDA to continue to provide a comprehensive evaluation of the portfolio of energy efficiency programs. Formal and informal feedback has indicated that the framework has achieved the goals for the evaluation established by the PSC.

References

Case 05-M-0900, In the Matter of the Systems Benefit Charge III, Order Continuing the Systems Benefit Charge (SBC) and the SBC-Funded Public Benefit Programs, issued and effective December 21, 2005.

DeCotis, Paul A., Mahone, Doug. 2005. "An Integrated Portfolio Approach to Evaluating Energy Programs: New York's Experience." *2005 International Energy Program Evaluation Conference*. Brooklyn, NY: IEPEC.

DeCotis, Paul A., Perry, Nancy W. 1999. "Innovations in Performance Management: Integration of Strategy and Planning." *Evaluation in Transition: Working in a Competitive Industry Environment*. Proceedings of the International Energy Program Evaluation Conference. Denver, Colorado.

NYSERDA (2006), "Facing Energy Challenges in the 21st Century, A Three-year Strategic Outlook", June 2006.

NYSERDA (2006). Systems Benefit Charge. *Proposed Plan for New York Energy $mart Programs* (2006 – 2011), as amended, March 2, 2006.

New York State Public Service Commission. In the Matter of Competitive Opportunities Regarding Electric Service, Opinion No. 98-3. *Opinion and Order Concerning System Benefits Charge Issues*. Issued and effective January 30, 1998. Cases 94-E-092 *et al*.

Order Continuing and Expanding the System Benefits Charge for Public Benefits Program. January 2001. New York State Public Service Commission.

Research into Action, Inc., "Review of the Evaluation Process", NYSERDA, February 2007.

Summit Blue, "ENERGY STAR® Labeled Homes and Home Performance with ENERGY STAR® Phase 1 Market Characterization, Assessment, and Causality (MCAC) Evaluation", NYSERDA, July 2004.