Large Lessons Learned: Impact Evaluation of Projects That Reported Over 1,500,000 kWh/yr Savings[1]

Jonathan B. Maxwell, ERS, College Station, TX
Cherie Gregoire and Jennifer Meissner, NYSERDA, Albany, NY
Lori Megdal, Ph.D., Megdal & Associates, LLC, Acton, MA

ABSTRACT

Evaluating projects that claim a large amount of savings presents unique challenges. This paper describes unconventional and successful techniques that an evaluation team used to compute gross realization rates and net-to-gross attribution for twenty-five large projects that together reported 128 million kilowatt-hours per year in gross electricity savings.[2]

Background

The New York State Energy Research and Development Authority (NYSERDA) and its evaluation contractors conducted a risk analysis in 2007 in order to help evaluators target limited impact evaluation resources for upcoming studies (Meissner, et al. 2008; Megdal & Associates 2007). The risk analysis identified sources of uncertainty among the factors that contribute to impact estimates across NYSERDA's portfolio of efficiency programs and quantified the uncertainty by factor.

The analysis revealed that the substantial contribution of NYSERDA's largest programs to the entire portfolio combined with relatively high uncertainty associated with the net-to-gross (NTG) estimates meant that these large programs contain a significant proportion of the overall uncertainty for the portfolio of programs. The largest projects in these programs can disproportionately affect the portfolio-wide results. In a portfolio of over three dozen programs and thousands of projects completed per year, the twenty-five largest projects in NYSERDA's portfolio account for more than 15 percent of the portfolio's savings installed in a typical year. Significant uncertainty can be eliminated by performing rigorous realization rate and NTG analysis on each of these projects.

The evaluated projects varied in size and technology. Projects ranged from 1.5 million to 10 million kWh/yr gross savings with implementation costs as much as $16 million. They included participants in five different NYSERDA programs. Technologies included:

- Combined heat and power (CHP) (5)
- Warehouse high-bay lighting (3)
- Comprehensive new construction (2)
- Steam cooling (1)
- Efficient snow guns (1)
- Low-flow fume hoods (1)
- Digester methane recovery (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] As of the time of writing this paper, only Phase I, 14 out of 25 sites, have been completed. All will be completed prior to the IEPEC 2009 conference and reported upon at that time.

It must be emphasized that these results are only being applied to the specific projects studied. As this is a census of the largest expected savers, results cannot be applied to any other groups of participants in any of these programs or used to provide any general conclusions about the programs themselves. However, the Large Savers study will inform the programs on methodologies for calculating savings estimates for specific technologies (*e.g.*, snow guns, specific CHP technologies, etc.) and increase the reliability of NYSERDA's overall program and portfolio *ex post* savings estimates due to the increased reliability in the estimates for these specific large projects.

Approach

Adjusted Gross Savings Calculation Approach

Each site was assigned a senior engineer as its lead based upon the site's technologies and its location and the expertise of the various senior engineers from among four engineering firms within the impact evaluation team. The lead impact evaluation engineers prepared measurement and verification (M&V) plans for each individual site. They reviewed the project files in detail, verified qualification for the study, and outlined the evaluation procedure. For multi-measure projects, the engineers performed Pareto analysis and analyzed only the measures with significant influence on the project overall impact. In all projects, more than 90 percent of the claimed savings was evaluated. In most of the projects, 100 percent of the claimed savings was evaluated.

The M&V plan for each site was based on an eight-page template specifying measure(s), approach, rigor, accuracy, data requirements, analysis, and budget for the evaluation.[3] The engineering director for the evaluation team and NYSERDA's impact evaluation manager reviewed and approved each M&V plan.

In forming the plan, the evaluation engineers had the option of proposing to base the analysis on simulation of building systems and equipment operation, analysis of utility bills, or direct measurement of equipment performance through metering or monitoring with appropriate instrumentation. Table 1 summarizes the approaches used for the sites. Most of the direct measurement sites required modeling to extrapolate short term metering to long-term annual performance estimates.

Table 1. M&V Approach Used

Parameter	Number of Projects
Utility bill data only (interval meters)	1
Direct measurement only	9
Utility meters and direct measurement	2
Building simulation modeling only	0
Other	1

[3] The plan template was drafted in November 2007, about the same time the CPUC-sponsored Evaluation Engineering Working Group committee was developing a similar M&V plan template for California 2006-2008 evaluation work. The NYSERDA template's primary author was part of the CPUC committee. While not identical, the two templates informed each other's development.

Due to their size many projects already had been subject to extensive measurement in response to customer or incentive program requirements. Evaluation engineers had extensive license to validate and then make use of this data, which ranged from HVAC energy management system downloads to twenty-four-month logs of compressed air and water flow rates for snow guns. Such data was especially useful when estimating pre-retrofit energy use for efficiency projects and for estimating post-retrofit energy use for CHP projects. In every case the evaluation engineer's method and execution was independent of any previously used approach and intended to add value to past measurement. Half of the site evaluations including those for all of the distributed generation projects used applicant-collected data.

Data collection typically involved an initial site visit with multiple interviews, logger installation, and spot measurement. A second visit followed two to four weeks later to collect loggers and ask follow-up questions.

After completion of the data gathering, the lead evaluator analyzed the data and reported the results in a standardized format. For this study of only twenty-five sites, this format was an Excel template and a Word narrative response to a set of questions. The engineering director reviewed all site-specific analysis and reports, wrote a half-page summary report for each evaluation, and performed the summary realization rate analysis.

Early in the analysis cycle evaluators determined that one site was a complete free rider, that is, the customer would have implemented the project even if the NYSERDA program did not exist. Realization rate analysis on that site was abandoned to allow for better investment elsewhere. At another site it became clear early that the site had a gross savings realization rate of 0; thus, attribution analysis was abandoned at this site. The overall guiding philosophy for data collection was to maximize the improvement in the savings estimate per evaluation dollar spent. The fact that the study results for these evaluated sites were not used to adjust the impact of other projects made this approach particularly justified, but with some care the method could be applied in other sampling scenarios as well.

Net-to-Gross Approach

NYSERDA defines the net impact factors as free ridership, participant on-site spillover, participant off-site spillover, and non-participant spillover. Spillover is the extent to which participation in the program inspired implementation of other energy-saving projects that were not funded by NYSERDA. The primary net-to-gross study objective was to develop site-specific estimates for the first three of these four factors. Non-participant spillover was the subject of a NYSERDA separate study incorporated into this large savers analysis as a program level, as opposed to site-specific attribution factor.

The secondary objective was to test multiple attribution methodologies simultaneously on the same sites. Researchers compared results as additional layers of evaluation complexity were added to the approach to test construct validity for the inquiries to measure the underlying construct of free ridership. (See Megdal, et al. 2009 for further detail on the free-ridership component of the Large Savers evaluation.)

Even the most basic approach was complex compared to some net-to-gross evaluations. Building on telephone survey instruments developed with NYSERDA by another evaluation contractor in prior studies,[4] the Megdal & Associates team created basic questionnaires to estimate attribution based on self-reporting, with the following characteristics:

- Multiple free-ridership queries per respondent

[4] Summit Blue Consulting.

    - Direct query
    - Likelihood Likert scale questions
    - Level of influence
    - Acceleration of installation
- Customized questionnaires for each NYSERDA program
- Inquiries on firm/corporate decision-making practices for various equipment/process improvements
    - Inquiry on firm criteria based upon type of decision maker
- Assessment of all of the above based upon both qualitative and quantitative factors

This new approach added several layers of complexity to the original direct query instrument. The first was to expand the target interview audience to include all possible decision makers. Vendors, project developers, and firm members of decision committees or within chains of approval were candidates for this influence pool. This aspect required that the interviews start with a series of decision-maker questions to identify all influencers.

The second level of complexity was that the lead engineer edited each site's set of attribution questionnaires for the particular program, technology, and respondent. The evaluation attribution lead and NYSERDA evaluation program manager reviewed and approved each site's set of questionnaires.

The final element of added complexity lay in the evaluation of the data. The lead engineer, a separate senior engineer not otherwise involved in the site's evaluation, and the attribution director each independently reviewed the completed questionnaires and estimated free ridership and participant spillover based on their own analysis of the survey question responses, that firms' decision-making process, and quotes from the interviews. The three then held a conference call to answer questions for the outside reviews and to discuss that firm's decision-making context and the responses obtained from each interviewee. The products of the call were a consensus range for that project's free ridership, a consensus final free-ridership point estimate, participant on-site spillover, and participant off-site spillover factors. The results presented later in this paper illustrate the level of variation in results for the differing methods. The uncommon amount of site-specific preparation and analysis resulted in highly informed attribution estimates based on the multiple viewpoints within a corporate decision-making process for each project.

Customer Recruitment and Interviewing

Large savings projects were treated as a census stratum: a customer that refuses to cooperate is irreplaceable. The participants are prominent customers, including leaders in manufacturing, communications, and health care. Evaluators used several techniques to respect the value of each customer to the analysis and their importance to the program administrators. All contacts were made by one person, the senior engineer in charge of the project who understood the technicalities of the project. The project did not rely on computer-assisted telephone interview (CATI) center professionals for any of the recruitment, scheduling, or interviews.

The site contact typically was only called after significant preparation and project-specific discussion with NYSERDA personnel. The engineer contacted the NYSERDA implementation program or project manager for each site first to give staff the opportunity to provide technical background on the project, offer contact advice, and otherwise prepare the engineer for the interviews. The NYSERDA project manager also sometimes helped the evaluators gain access to the participants.

The second set of contacts targeted was the third-party developer or consultant. Again the goal of using this order was to maximize background preparation prior to talking with the site contact. Third parties were also often most familiar with past analysis and available data. The engineers typically completed developer attribution interviews during these phone calls.

The third and final contact was with the site participants. All participant interviews were conducted in person by the lead engineer, unless a decision maker was out of state or insisted on telephone-based interviews for their convenience. The same senior engineer led all data collection. ERS staff had experience with such an approach in highly technical interview situations (Maxwell, et al. 2001) and other evaluation professionals have supported such an approach (Goldberg and Scheuermann 1997).

The sequential nature of interviewing first program staff, then developers, then host facilities required more calendar time than simply calling the site directly, but the evaluation administrators found the extra time worthwhile because it enabled them to treat important customers with maximum care and resulted in identification of excellent data sources for some of the projects.

Pursuit of interviews with key decision makers and securing site and/or data access was allowed to continue for as long as nine months. This schedule is not possible with all evaluations. In this project it enabled evaluators to complete over 90 percent of the target evaluations in an all-volunteer basis. In certain cases the extra time gave participants a chance to implement projects identified in NYSERDA co-funded studies or to collect needed data.

In what may be seen as an unconventional approach, the evaluation directors gave the senior engineers the latitude to conduct the interviews in conversational fashion as opposed to reading the script. Though the scripts were predominantly multiple choice, interviewers were taught to engage in open-ended discussions and later complete the questionnaires as appropriate. Such techniques were expensive and necessitated training on proper interviewing techniques (*e.g.*, no leading questions) and also required senior staff. But they were also respondent-friendly and enabled superior interview data quality in the form of nuanced understanding of decision-making dynamics by individuals intimately familiar with complex technologies.

General Comments on Approach

This is a census of the largest expected savers. Results were not applied to any other groups of participants in any of these NYSERDA programs and do not provide any general conclusions about the programs overall. The project increased the reliability of NYSERDA's overall program and portfolio estimates due to the increased reliability in the estimates for these specific large projects.

Results

Adjusted Gross Impact Results

Figure 1 and Table 2 summarize the realization rates (evaluated savings divided by program claimed savings) for each of the evaluated projects.

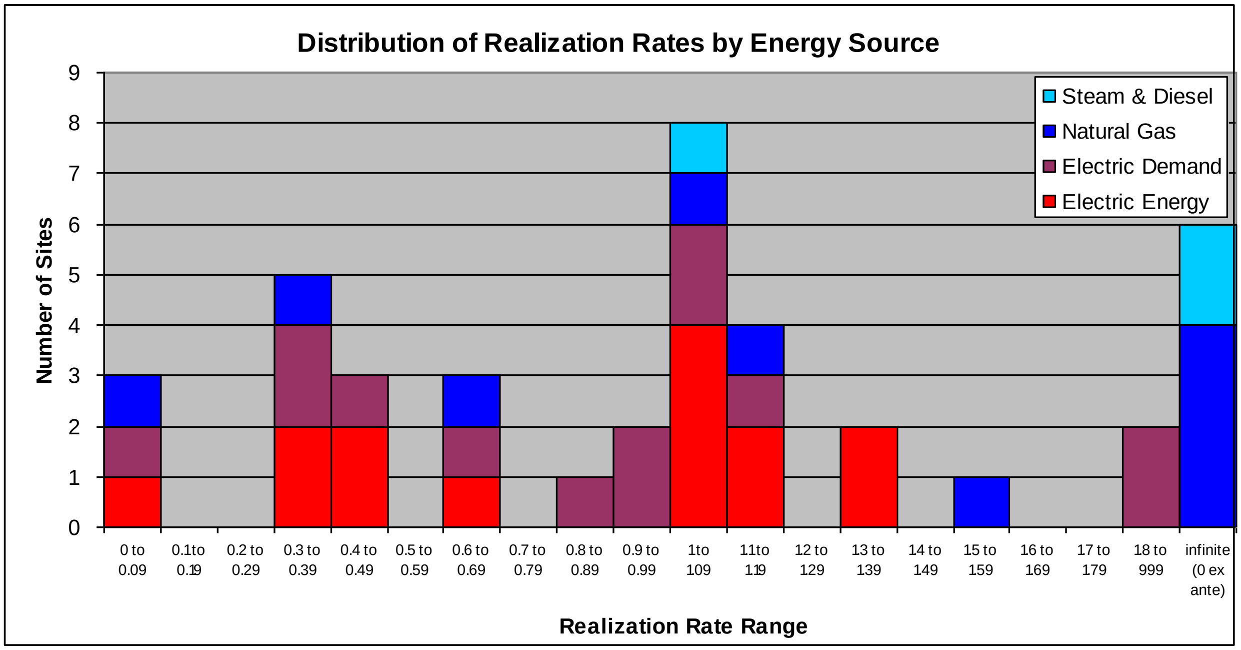


Figure 1. Adjusted Gross Realization Rates by Site

Table 2. Adjusted Gross Realization Rate by Program

Program	kWh	kW	Gas	Steam	Diesel
R&D DG-CHP	91%	101%	83%		na
Technical Assistance	39%	34%			
Peak Load Mgmt	107%	107%	174%	107%	
C/I Performance	91%	91%	na		na
New Construction	61%	61%	98%		
Total	**82%**	**82%**	**101%**		**na**

The Technical Assistance (TA) project realization rate encompasses both the measure adoption realization rate and the gross savings realization rate. That is, the denominator is the maximum savings the customer could have realized by maximally implementing all measures in the study. A realization rate of less than 100 percent does not necessarily indicate an installed measure is saving less than expected within the TA program. For example, if a TA participant installs half of the recommended measures and has a gross savings realization rate on those measures of 100 percent, then the overall gross savings realization rate for that participant is 50 percent due to the installation rate.

The program-reported savings values that are the denominators of the realization rates (e.g., *ex ante* impacts) are not necessarily what the vendors promised or what the participants expected at the time the decision to proceed was made. They are what NYSERDA's program administrators estimated prior to evaluation, which sometimes differed from vendor or customer estimates. For example:

- In multiple instances the *ex ante* impact associated with the project at the time of funding approval was less than promised in application literature. NYSERDA reduced the impact reported by the program due to past experience with the technology. This increased the evaluated realization rate and was responsible management, but perhaps also veiled some lost savings potential.
- In multiple cases the *ex ante* impacts were based on extensive (up to two years) post-installation metering by the participant or developer. In those cases the *ex ante* values already were adjusted from pre-installation estimates. Even the use of such extensive post-retrofit data collection did not assure a realization rate near 100 percent. One of the projects with such intensive post-retrofit study had a realization rate of less than 50 percent.

The study's overall realization rates were in the range found in many impact evaluations—82 percent to 107 percent, depending on fuel. The more revealing statistic is that for these projects—huge investments that required sometimes hundreds of thousands of dollars in pre-decision research and tens of thousands of dollars in post-installation metering—the average evaluated savings (*ex post* kWh/yr, kW, or MMBtu/yr impact) differed from *ex ante* impact by 42 percent. Excluding the one project with a 0 realization rate decreases the average deviation to 37 percent. On the other hand, this average doesn't account for the six instances where reported savings for a fuel was 0 and evaluated savings was non-zero. All of the lighting projects had realization rates closer to 100 percent than the average. The average difference was 48 percent without lighting. Table 3 summarizes the standard deviations of the realization rates. The graphic presentation of the large variation in gross savings realization rates is easily seen by re-examining Figure 1 and the dispersion of the gross savings realization rates across these large and complicated projects.

Table 3. Standard Deviation of Realization Rates by Fuel

Parameter	Electricity	Peak Demand	Gas
Standard deviation on realization rate	46%	66%	60%

In short, variation was high. It appears that the large scale, uniqueness, and complexities of these projects trumped the additional *ex ante* analysis performed to determine savings. In retrospect the actual engineering uncertainty of these projects was much higher than the ±10 percent often projected by auditors, vendors, and third-party M&V plans, probably because they were so unusual. Despite the high variation, the overall average realization rate is within the range of traditional expectations.

For individuals responsible for writing the rules of program evaluation, the variation in realization rates even while using substantial amounts of program information demonstrates an important methodological lesson. Allowing evaluators to use program-generated data and results does not necessarily result in convergence on 1.0 realization rates. Rigorous evaluation engineering can leverage program data without necessarily resulting in the same or biased estimates.

Net-to-Gross Results—Free Ridership

The free-ridership level seen in Phase I projects is fairly high. Program-level results are included in Table 4.[5] This result is not completely unexpected given a higher proportion of the largest customers can be expected to have well-trained engineering staff and internal resources to search, consider, and finance efficiency improvements.

Free ridership is based upon customer knowledge concerning the equipment and building options, the vendor's depth of experience and knowledge of efficient equipment, the customer's decision-making process and financial situation, and the circumstances surrounding the equipment purchase or building construction. There can be significant variation across customers and projects.

Table 4. Phase I Project Free-Ridership Estimates

		Free-Ridership Estimate[a]		
Project	**n**	**Prior Method**	**Lead Engineer Estimate Prior to Assessment Conference**	**Final Consensus Estimate**
Peak Load Management	3	69%	53%	59%
Commercial/Industrial Process	3	7%	3%	10%
New Construction	3	85%	89%	88%
Distributed Generation--Combined Heat & Power	4	59%	65%	52%
All Programs	**14**	**53%**		**53%**

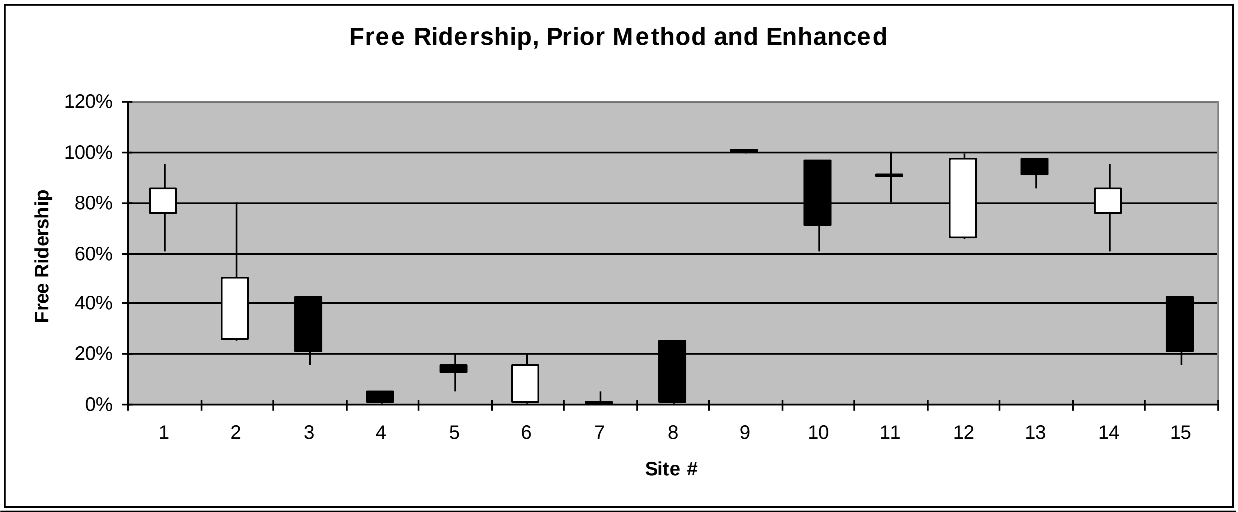
[a] All estimates are weighted by the *ex ante* savings estimates.

These results are for the projects studied only and will not be applied to the entire program. The program name is used so as to maintain the confidentiality of the participants in the study.

Figure 2 compares the free-ridership estimates using the two different methods for individual sites.[6] The graph illustrates much the same point as the realization rate. Despite the similarities in the overall average values, there were significant differences at the site level. The average adjustment was 14 percent. For over half of the sites the prior method estimate was outside of the consensus range.

[5] The TA project is not within these program-by-program free-ridership results so as to maintain the confidentiality of the participants.

[6] The site numbers in this chart have been intentionally reordered and do not correspond with the numbers in other charts.



Legend: The thin vertical line for each site represents the consensus-based (enhanced) free ridership ranges. The thicker bars represent the movement between the prior method estimate and the consensus-based point estimate. If the bar is white, then the prior estimate was the low end of the bar and the consensus-based point estimate is the high end of the bar. If the bar is black, then the prior method estimate was higher and the consensus-based point estimate was lower. *If either end of a thin vertical line is not visible, the prior method estimate for the site was completely outside of the consensus-based estimate range.*

Figure 2. Free-Ridership Estimate Comparisons and Ranges

Net-to-Gross Results—Participant Spillover

Table 5 summarizes the spillover findings. The values are low relative to free ridership. Savings generated, however, are not trivial given the large size of these projects.

Table 5. Phase I Project Free-Ridership Estimates

Project	n	Spillover Estimate[a]		
		Participant On-Site	Participant Off-Site	Developer
Peak Load Management	3	0%	5%	0%
Commercial/Industrial Process	3	1%	4%	0%
New Construction	3	0%	0%	0%
Distributed Generation--Combined Heat & Power	4	0%	0%	0%

[a] All estimates are weighted by the *ex ante* savings estimates.

These results are for the projects studied only, and will not be applied to the entire program. The program name is used so as to maintain the confidentiality of the participants in the study.

Evaluation Economics

A recent ERS survey of energy efficiency programs around the country found that sponsors typically fund implementation at levels ranging between $0.12 to $0.24 per annual kWh saved. Evaluation funding typically is 2% to 8% of implementation funding. Using the midpoints of those ranges, typical evaluation funding is about $0.009 (5% of $0.18) per gross annual kWh saved.

Total funding for this large savers evaluation was about $330,000. Based on total evaluated energy savings of 128 million kWh/year, the evaluation cost was $0.0025 /kWh/yr saved. This means that the Large Savers evaluation study was completed for 70 percent less than a typical evaluation and without sampling error.

Conclusions

Evaluation of large energy saving projects entails different challenges and techniques than evaluation of mass market programs. For this project, evaluators successfully used the following techniques to maximize accuracy and research value:

- Required site-specific M&V plan for each project
- Administered all net-to-gross survey instruments by senior engineers with significant input from a senior evaluator with social science training and experience
- Allowed evaluators to use program-collected data
- Allowed use of IPMVP Option A, B, C, or D level analysis or hybrid approaches depending on the nature of the project
- Planned for extended calendar time to complete the study
- Developed site-specific survey instruments for gross, net-to-gross, and non-energy impact inquiries
- Encouraged open-ended discussion type interviews
- Interviewed multiple decision makers per site
- Developed consensus free-ridership and spillover estimates with multiple reviewers of attribution data and extensive discussions of both quantitative and qualitative information collected and observed through the evaluation process

The evaluation found substantial variation in realization rates. Even with the significant effort already invested by participants and developers in estimating savings, the *ex post* impact estimates differed from *ex ante* estimates by an average of 42 percent. Overall the weighted average realization rates were 82 percent on electric energy, 86 percent on peak demand, and 102 percent on natural gas. Realization rates would have been lower if NYSERDA program staff had not already anticipated and adjusted for system underperformance for certain technologies.

Attribution evaluation generally found high free ridership for these large savers, 53 percent overall. Participant spillover varied from 0 percent to 5 percent depending on the program. These findings, however, hide substantial variation by customer type, technology, and decision-making process. There are large institutions where knowledgeable facility engineers propose and “intend” to invest in energy efficiency but where governmental or institutional processes hinder final proposal or financing at the final decision-making level. There are many large entities with large sophisticated engineering staffs where these investments are being made. Yet, there are also technologies that are new and without program promotion, analysis, or market preparation will not otherwise be adopted.

NYSERDA's Large Savers' evaluation study developed and conducted in-depth customer and technology specific gross and net evaluation methods, instruments, and analyses. This in-depth evaluation was also cost-effective. It cost 70 percent less than typical evaluations on a dollars per annual kWh saved basis, in spite of an extended M&V schedule. At the same time, the results were delivered with 0 percent sampling error for the subject population.

References

Goldberg, Miriam L. and Kurt Scheuermann. 1997. "Gross and Net Savings for Unique Projects." *In Proceedings of the 1997 International Energy Program Evaluation Conference*, Chicago, IL: International Energy Program Evaluation Conference.

Maxwell, Jonathan B., Roland Clarke, Fred Gordon, Edward McGlynn, Hale Powell. 2001. "Benchmarking Current System-Wide Compressed Air Market Efficiency Practices in the Northeast United States and Programmatic Strategies to Improve Them." *In Proceedings of the 2001 ACEEE Summer Study on Industry*. Tarrytown, NY: American Council for an Energy-Efficiency Economy.

Megdal & Associates. 2007. Risk Analysis for NYSERDA's System Benefit Charge (SBC)-Funded Impact Evaluation Estimates, Prepared for the New York State Energy Research and Development Authority (NYSERDA), Principal investigators: Steven Meyers of Rational Energy Network, Dr. Lori Megdal, Megdal & Associates, and Kathryn Parlin, West Hill Energy & Computing, November.

Megdal, Lori M., Yogesh Patil, Cherie Gregoire, Jennifer Meissner, and Kathryn Parlin. 2009. "Feasting at the Ultimate Enhanced Free-Ridership Salad Bar." *In Proceedings of the 2009 International Energy Program Evaluation Conference*, Portland, OR: International Energy Program Evaluation Conference.

Meissner, Jennifer, Cherie Gregoire, Steven Meyers, Lori Megdal, and Kathryn Parlin. 2008. "Allocating Impact Evaluation Resources: Using Risk Analysis to get the Biggest Bang for your Buck." *In Proceedings of the 18th National Energy Services Conference*, Clearwater Beach, FL: Association of Energy Services Professionals.