The Challenges of Conducting an Impact Evaluation in Real-Time[1]

Rebecca Reed, NYSERDA, Albany, NY
Jennifer Meissner, NYSERDA, Albany, NY
Shawn Shaw, Cadmus, Waltham, MA

ABSTRACT

The American Reinvestment and Recovery Act (ARRA) was signed into law in February 2009. The New York State Energy Research and Development Authority (NYSERDA) received $171.6 million of ARRA funds to offer the residents of New York State a series of energy efficiency and renewable energy programs. Working with an independent consultant, NYSERDA conducted an impact evaluation to determine the amount of attributable energy and demand savings, renewable energy generation and capacity, carbon emissions reductions and job creation of the ARRA-funded programs. The evaluation also examined program cost-effectiveness. A key provision of ARRA was to spend the funds and conduct the work, including evaluation, within the Program Performance Period. This paper will discuss the benefits and drawbacks of conducting an impact evaluation in real-time and the challenges associated with reporting on a program's impact before the program is completed.

BACKGROUND

On February 13, 2009, in response to a deepening recession in the United States economy, the U.S. Congress passed the American Reinvestment and Recovery Act (ARRA). The legislation was signed into law by President Obama on February 17, 2009. ARRA expressed three immediate goals:

1. To create new jobs, as well as save existing ones,
2. To spur economic activity and invest in long-term economic growth, and
3. To foster unprecedented levels of accountability and transparency in government spending.

The intention of ARRA was to fund shovel-ready projects that could begin construction as soon as possible.

Components of this law made funding available to states through two separate programs managed by the U.S. Department of Energy (DOE). These included the State Energy Program (SEP) and the Energy Efficiency and Conservation Block Grant (EECBG). NYSERDA received a total of $152.9 million in funding through these two programs ($123.1 million from SEP, and $29.8 million from EECBG), which was combined with $18.7 million in State Energy Efficiency Appliance Rebate Program (SEEARP) funding, also received through ARRA. Collectively, these funds were offered to public and private sector organizations and residents of the State of New York through energy-efficiency, renewable generation and transportation incentive programs, energy code training programs and energy conservation study opportunities.

To streamline the evaluation work, the ARRA programs were organized and consolidated into six distinct Program Areas: Appliance Rebate, Energy Code, Energy Conservation Studies, Energy-Efficiency, Renewable Energy, and Transportation. Table 1 presents a summary of the ARRA-funded Program Areas

[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.

that were included in the evaluation. Renewable Energy and Energy-Efficiency Program Areas received the majority of the funding.

Table 1: ARRA-Funded Program Areas and Expenditures

Program Area	ARRA Funding Source	Total Projected Program Area Expenditures ($ million)*
Renewable Energy	SEP, EECBG	$46.7
Energy-Efficiency	SEP, EECBG	$52.9
Transportation	SEP, EECBG	$3.8
Energy Code	SEP, EECBG	$6.6
Energy Conservation Studies	SEP	$5.3
Appliance Rebate	SEEARP	$18.7
TOTAL	ALL	$134.0

* Program Areas and expenditures presented in this table do not include ARRA-funded programs that were not a part of the evaluation, including Material Conservation, Energy Management Personnel, subsequent rounds of the Appliance Rebate Program, and Energy Code Compliance.

These ARRA-funded Program Areas were designed to be unique from, but complement, NYSERDA's existing robust and diversified portfolio of energy-efficiency and renewable energy programs. NYSERDA's energy-efficiency and renewable energy programs support the policy goal of meeting 45% of the State's electricity needs through improved energy efficiency and clean renewable energy by the year 2015 (45 x 15 goal).[2]

In 2009, NYSERDA issued a Request for Proposals (RFP) for evaluation services to determine the impacts of these ARRA-funded Program Areas. The contract was awarded to a team (the Team) led by The Cadmus Group (Cadmus). The primary objectives of the evaluation included:

- Determining attributable energy and demand savings by Program Area
- Quantifying renewable energy capacity and generation attributable to the Renewable Energy Program Area
- Computing the displacement of greenhouse gas (GHG) emissions and quantifying the environmental impacts associated with each Program Area
- Evaluating the economic impacts (including job creation and retention) created by all Program Areas
- Determining the cost-effectiveness of ARRA-funded Program Areas

The Team achieved the objectives above through file reviews, site visits and surveys. Many of the site visits included monitoring and verification (M&V) efforts. The Team also assessed customer satisfaction where possible by including additional customer satisfaction questions to the survey instruments used to address the objectives above.

[2] On June 28, 2008, the New York State Public Service Commission (PSC) issued an Order approving the Energy Efficiency Portfolio Standard programs to reduce energy consumption in New York State by a total of 15% below the 2006 forecast for the year 2015; referred to as the 15x15 goal. On January 8, 2010, the PSC expanded the Renewable Portfolio Standard goal to increase the proportion of renewable electricity consumed by New York customers from 25% to 30% and extended the terminal year of the program from 2013 to 2015. Collectively, these two goals are referred to as 45 x 15.

To the greatest extent possible, the Team ensured that the work undertaken for this evaluation was completed in accordance with the evaluation guidelines[3] established by the DOE for ARRA-funded programs, and with New York's existing evaluation guidelines for ratepayer-funded energy-efficiency programs.[4] The Team evaluated the outcomes by Program Area, at the portfolio level, and by funding source. First, the Team created sample designs for each technology grouping - or Program Area - under each of the funding streams, with a goal of achieving 90/10 sampling precision on gross energy savings/generation for the overall funding source. Additionally, the Team examined the portfolio of Program Areas as a whole. Lastly, the Team considered activities funded through each of the major ARRA-funding streams (SEP, EECBG, and SEEARP).

At the same time as the evaluation reports were written, many of NYSERDA's ARRA-funded Program Areas were continuing to operate, and many of the planned projects to which funds had been committed had not yet been completed. Due to DOE requirements, and the contract between Cadmus and NYSERDA, the evaluation reports for these Program Areas were due to be completed by the end of the original Program Performance Periods, March 31, 2012 and September 30, 2012 for SEP and EECBG respectively. Thus, evaluation of programs had to be completed sometimes before the programs were actually complete. Evaluating programs before they are complete has benefits and drawbacks related to project sampling, site work, attribution assessment, and analysis and reporting. Therefore, NYSERDA's ARRA experience provides valuable insight as a case study to identify some important benefits and drawbacks of real-time evaluation.

Project Sampling

To evaluate the impacts of the ARRA-funded programs, the Team conducted numerous surveys (i.e., phone, on-line or in-person surveys) and site visits for a sample of projects that, ideally, represent the same characteristics as the final population of ARRA-funded projects. As required by DOE evaluation guidelines, all sampling must meet or exceed a 10% margin of error at the 90% confidence level for the overall funding source. Likewise, the New York State evaluation guidelines for ratepayer funded programs have a similar requirement. For the Renewable Energy Program Area, the site visit sample was drawn to meet the 90/10 confidence and precision level. Each customer in the site visit sample also completed a survey either in person or on-line. Additionally, another 10% of the Renewable Energy projects completed an on-line survey. Table 2 shows the number of projects funded by ARRA, sites visited and surveys conducted by Program Area.

[3] Guidance for EECBG recipients: http://www1.eere.energy.gov/wip/pdfs/eecbg_evaluation_guidelines_10_017.pdf; Guidance for SEP recipients: http://www.tecmarket.net/documents/Final%20SEP%20Evaluation%20White%20Paper%2010-18.pdf

[4] Guidance for EEPS Evaluation: http://www.dps.ny.gov/EEPS_Evaluation.html

Table 2: Site visit and survey sample sizes by Program Area

Program Area	Number of Projects	Number of Surveys Completed	Number of Site Visits Conducted
Renewable Energy	217	95	77
Energy-Efficiency	188	65	87
Transportation	9	9	0
Energy Code	N/A	1,256	0
Energy Conservation Studies	204	42	0
Appliance Rebate	169,866	560	0
TOTAL	170,484	2,027	164

In order to allow for sufficient time to contact program recipients, conduct surveys and/or site visits, analyze results, and summarize findings in a detailed report, the sample for these ARRA-funded projects needed to be drawn several months prior to the end of the Program Performance Period. Determining the population from which to draw a sample proved challenging for several reasons:

1. Most projects were not completed by the time sampling and data collection needed to begin,
2. Many projects went through multiple scope changes, and
3. Estimated project completion dates changed frequently.

To address and overcome these obstacles, the Team took one of two approaches for each Program Area:

1. Projects were surveyed and/or visited as they were completed; or
2. A population was constructed using a specific cut-off date based on estimated project completion dates, and a sample was drawn from this population.

The first approach of visiting or surveying projects as they were completed (in effect a rolling sample strategy) was used for Program Areas that had a small population with relatively few projects completed at the time the sample needed to be drawn. Program Areas that used a rolling sample strategy included Transportation and Renewable Energy. In many cases, these rolling samples were ultimately census attempts. A census attempt is when every project in the population is contacted for the survey or site visit, so no sampling is required. While this approach led to a lengthened evaluation period and late data collection (some occurring while the initial draft evaluation report was being written), there were obvious benefits in terms of representativeness of the evaluated projects, since an attempt was made to survey or visit every single project. Using this rolling sampling strategy was beneficial for monitoring and verification (M&V) efforts, as it enabled the Team to begin collecting on-site metering data soon after an individual measure was installed, often providing months of data. Had the M&V sample been drawn near the end of the process, only a short period of data would have been available for M&V, even on systems installed earlier in the Program. The main drawback of this approach is that it required The Team to spend extra time staying abreast of project construction and implementation schedules so that metering equipment could be cost-effectively incorporated into the system design, if needed.

The second approach determined the population and sample by using a specific project installation cut-off date. This approach was used for Program Areas with a larger population that had many projects. In some cases a large proportion of the total expected projects were already completed. These Program Areas included the Appliance Rebate, Energy Conservation Studies, and Energy Efficiency. A significant sampling concern with this second method was whether the sample would be representative of the final population.

This was of particular concern with Program Areas falling into the second approach category since site visits and surveys were only conducted for projects that were installed months before the conclusion of the Program Performance Period. Fortunately, ex post analysis showed that this was not an issue since the mix of technology and project types used in the samples were quite similar to the final mix of projects awarded funding under ARRA. One benefit of this sampling approach was that the Team could start and finish their field work in a reasonably predictable timeframe. This allowed for ample time to conduct analyses and resulted in less modification of outcomes as the draft report was being written.

Nevertheless, the Team had to remain flexible with both sampling approaches, and many times replacement sites were needed for projects that were delayed or canceled. This required extra effort, on the part of both program implementers and evaluators. This extra effort included scheduling coordination of sites for analysis, and ensuring sufficient evaluation resources were put toward projects that could be reliably sampled.

Conducting On-Site Work

The Team conducted over 160 site visits to collect measurements to determine actual energy savings or generation. Conducting the on-site work during or immediately after project implementation has several advantages. Foremost, it allows for the opportunity to coordinate site visits with the project implementers to help reduce the burden of external visitors and interactions on the facility manager. Additionally, in some cases, early coordination with project implementers allowed for monitoring equipment to be installed during construction. While it is possible to conduct M&V after the measure is installed, some equipment, such as flow meters to measure the amount of heated water produced by a solar hot water system, can be expensive to install after the measure is operational. Installing a typical flow meter in a solar hot water system requires shutting the system down, cutting out a section of pipe, and installing the flow meter – a prohibitively disruptive task for a temporary M&V effort. Incorporating the flow meters into the system design and installation allowed the Team to collect very high quality data that otherwise would have been impossible, or prohibitively expensive, to collect.

Moreover, conducting the evaluation work in real-time provides more timely feedback to Program Staff regarding whether or not the program is on track to meet overall participation, spending, and energy savings goals, and whether or not the projects are achieving the expected energy savings. Early feedback on the effectiveness and barriers to program processes and program participant satisfaction can help Program Staff modify the program design as necessary to help achieve participation, spending and energy saving targets. Additionally, if monitoring and verification work identifies issues with project performance, these issues can be corrected sooner rather than later, both to the benefit of the customer and program, by increasing the energy-saving potential of projects.

For example, the Team monitored and collected data from two of the three ARRA-funded Solar Wall™ projects (see Figure 1)[5]. After the Team analyzed the Solar Wall data, the findings were presented to NYSERDA and the customer, and a recommendation was made by the Team to adjust the settings on one of the Solar Walls in order to optimize control settings to improve energy savings.

---

[5] The Solar Wall™, developed by Conserval Engineering, is a transpired solar collector used to preheat the fresh air intake of commercial and industrial buildings. For more information, see the manufacturer website: www.solarwall.com

Figure 1: The Team Deployed Equipment to Collect a Variety of Performance and Environmental Data on Solar Wall Performance

To conclude that adjustments were needed to the control settings on one of the ARRA Solar Wall projects in order to maximize the energy savings, the Team measured and calculated several parameters at the project site. First, the daily thermal energy contribution of the solar wall was calculated. The energy savings of the solar wall system was then considered; energy savings came primarily from offsetting consumption of two fan units at the site. The Team calculated the average energy consumption per day from the fans for all days with metered data and then scaled that to an annual value. Similarly, the Team correlated the daily thermal energy of the solar wall with daily Heating Degree Days (HDD) and incident solar radiation. Once this correlation was established for the monitoring period, the correlation terms were applied to Typical Meteorological Year (TMY), daily average irradiance, and HDD values.

In cases where the regression method resulted in a poor fit with observed data, a ratio-based method[6] was used to adjust monitored data to typical HDD and irradiance conditions. Furthermore, when considering energy savings, it was critically important to bear in mind that the air in a large open heated area will naturally stratify based on the laws of thermodynamics. This will cause warm air to rise and cold air to sink. Heating load and system performance could thus be impacted by this condition; stratification of the heated air could negatively impact system performance and heating demand. To address this potential problem, the Team reviewed vertical temperature profiles within the heated space to identify the amount of time with and without the solar wall operating and found that supplementing the solar wall with minimal gas heating could destratify the interior airspace and provide additional, unrealized generation. Thus, by conducting a detailed and thorough investigation for a very complex system, the results from monitoring in real-time were used to make recommendations to the customer to adjust the control settings on one of the solar wall units in order to maximize the effectiveness of the solar wall project.

Although there are benefits to real-time M&V site visits, coordination between the project implementers and evaluators is not always possible. When coordination is not possible, the real-time evaluation on-site work can be a major burden and even confusing to homeowners and facility mangers, who may be receiving multiple requests for site access during a short span of time. This was a significant and real issue for the ARRA-funded projects, which had a high administrative burden due to the frequent necessity of several site visits from various entities. All ARRA-funded projects were subject to at least one quality assurance/quality control (QA/QC) inspection by a third party contractor hired by NYSERDA to verify that the project was functional, installed as designed, and Buy-American and Davis-Bacon Act (DBA) compliant[7]. Complex and multi-phase projects often received multiple QA/QC inspections. Additionally,

[6] The ratio-based method compares the average solar irradiance and total HDDs during the monitoring period with typical values for the solar wall operating period, and applies this ratio to the observed savings number.

[7] Wage and Hour Division Information Related to the American Recovery and Reinvestment Act of 2009:

approximately 20 percent of the ARRA-funded projects were subject to site monitoring by a DOE Project Officer, and approximately ten percent of the ARRA-funded projects were subject to a site visit by the U.S. DOE Inspector General's Office. For example, a large New York City hospital facility received three QA/QC inspections from NYSERDA's third party contractor, two site visits from the DOE Project Officer, one site visit from the Inspector General's Office and two site visits from the Evaluation Team.

Finally, evaluating a newly-installed system or project doesn't allow sufficient time for a measure "shake down" period, and persistence cannot be measured. Conducting measurement and verification of equipment before the shake down period has passed can cause results to be biased by lower than expected savings caused by short-term issues with the measure. In the case of Solar Walls™, for example, the controls can be too complex to integrate into the building's existing HVAC system and may require several weeks of observation to optimize. Collecting data only during these initial weeks would lead to estimating unrealistically low energy savings for the project that are not representative of the system's long-term benefits.

Attribution

DOE evaluation guidelines for ARRA-funded programs required that the evaluation work should "[d]ocument the resulting effects (job creation/retention, energy savings, renewable energy generation, and carbon reductions) that are above and beyond the effects that would have been achieved without those funds. That is, studies should focus on net effects of the …Recovery Act initiatives."[8] These effects are referred to as attribution effects, and to capture them attribution surveys were used. Specifically, to determine program attribution, or those effects that are above and beyond what would have happened in the absence of the program funding, the Team conducted either phone, on-line or in-person surveys with almost 800 decision-makers across the portfolios and Program Areas.

The Team assessed attribution to determine what would have happened in the absence of the Program Areas. In essence, the Team measured a counterfactual: something that never actually happened. The nature of ARRA funding added another layer of complexity to attribution analysis. The federal government provided strong directives to award money to projects that were ready to move forward—known as being shovel ready—but that had difficulty securing financing due to the recession. The Team took such directives into account when assessing attribution, adjusting the definition of freeridership to account for the degree to which ARRA funds allowed projects to continue that might have been delayed or would have been scaled back without ARRA funds. Hence, the Team sought to determine:

1. Whether the energy savings would have occurred in the absence of the Program Area efforts and funding
2. Whether the *same level* of energy savings would have occurred in the absence of the Program Area efforts and funding
3. Whether the savings would have occurred at the *same time* in the absence of the Program Area efforts and funding

---

http://www.dol.gov/whd/recovery/. ARRA Buy American Provision: http://www1.eere.energy.gov/recovery/buy_american_provision.html.

[8] Guidance for EECBG recipients: http://www1.eere.energy.gov/wip/pdfs/eecbg_evaluation_guidelines_10_017.pdf, page 3. Guidance for SEP recipients: http://www.tecmarket.net/documents/Final%20SEP%20Evaluation%20White%20Paper%2010-18.pdf

In order to best compare Program Area attribution results with those of other NYSERDA programs, the Team attempted to follow the long-standing approaches for measuring attribution that have been used in other NYSERDA evaluations[9], as well as those considered best practices in the industry. These inquiries sought to determine the level of freeridership in the programs. Freeridership is the proportion of energy savings and renewable energy generation that would have occurred in absence of the incentives.

One of the greatest benefits of conducting evaluations close to the time of project completion, and while programs are actively operating, is that evaluators are able to speak with customers when they have recently gone through the decision-making process. The close proximity of evaluation to the decision making timeframe is expected to yield the greatest reliability in customer responses regarding the activities they would have been likely to undertake in the absence of a program. When evaluating completed programs or projects long since finished, evaluators are sometimes asking participants about decisions they made months or years earlier, and it is very difficult for customers to remember exactly how much influence a program may have had on their decision-making process. Additionally, due to turn over in the workforce, it is sometimes difficult for evaluators to find the decision-makers who were present at the time of the project when doing retrospective surveys.

In contrast, one of the greatest challenges to evaluating projects close to the time of completion or evaluating active and relatively new programs is that it is difficult or impossible to estimate the spillover impacts. For purposes of this evaluation, spillover is defined as reductions in energy consumption and/or demand caused by the presence of the energy-efficiency or renewable energy measures, beyond the direct energy benefit of the ARRA-funded measure itself. Spillover can be characterized as participant and/or non-participant spillover. Spillover may take months or years to occur depending on factors such as market maturity, program design, measures/technologies supported, cost, and experience a customer has with program measures. In most cases for this evaluation effort, the Team did not attempt to measure spillover because customers did not have sufficient time to pursue other actions or purchase equipment they may have become aware of through their participation in the Program Area.

To determine the amount of attributable energy and demand savings and renewable energy capacity and generation, the Team calculated two sets of numbers: Evaluated Savings/Generation and Projected Savings/Generation. The Evaluated Savings/Generation included projects that were installed or operational prior to December 31, 2011 for SEP-funded activities and prior to June 30, 2012 for EECBG-funded activities. The evaluated savings are limited to evaluated, measured, verified savings, and savings net of freeridership. Because not enough time had passed for the Team to assess spillover, the evaluated savings are adjusted only for freeridership, and likely understate the actual savings that may be realized.

The Projected Savings/Generation include the energy savings and generation for all projects that were either completed, under contract, and/or expected to be completed by the end of the Program Performance Period. The Projected Savings/Generation numbers were adjusted for both freeridership and spillover by using a deemed Net-to-Gross (NTG) value of 0.9. NYSERDA and the New York Department of Public Service have a precedent of accepting a deemed NTG of 0.9 for planning purposes. Therefore, the Projected Savings/Generation methodology produces a long-term projection of the total savings that are most likely to occur from the ARRA-funded programs by including all completed and planned ARRA-funded projects, and adjusting for both freeridership and spillover.

---

[9] http://www.nyserda.ny.gov/en/Program-Evaluation/NYE$-Evaluation-Contractor-Reports.aspx.

Analysis and Reporting

As one might expect, given the short evaluation time line and real-time approach, analysis and reporting were also challenging for the Team. The Team had to weigh timing considerations with two competing goals:

1. To capture the greatest amount of program and project activity in the sampling and data collection, and
2. To leave enough time to conduct careful analysis and complete reporting on the vast amount of data collected through this evaluation effort.

In many cases project level data and analysis was being completed as the first draft reports were being written and reviewed. This meant that reviews and discussions were based on preliminary findings that in some cases required modification when the final results were complete. Another challenge related to conducting the overarching analysis of cost-effectiveness, job creation, and emission reductions was that preliminary data had to be modified to derive the final results.

Lessons Learned for Real-Time Evaluation

Whether real-time evaluation is a deliberately selected approach or done out of necessity, it can be an effective tool to provide credible assessment of program results that serve program implementers' needs for current information. There are several lessons learned which can be useful for future real-time evaluations:

1. A high level of coordination is needed between program implementers and evaluators to make real-time evaluation a success. Program implementers need to be willing to allow the evaluation to operate alongside, and be integrated along with, their day-to-day activities, as needed. Evaluators need to stay in touch with program implementers and have to take responsibility for obtaining near constant program and project status updates in order to most effectively deploy their resources. All of this must be done while maintaining the independence of the evaluation.
2. Evaluators must be flexible and willing to stretch beyond their normal comfort zone in terms of data collection, analysis and reporting periods. Often these periods of an evaluation are more distinct, but in real-time evaluation they may flow together and require effective management to work. Similarly, clients of evaluation need to be willing to accept preliminary results that could be subject to change as final results come in from the final surveys and site visits.
3. It is difficult, if not impossible, to accurately assess spillover impacts of a program when the evaluation is conducted concurrently with program implementation. Adjusting energy savings for only freeridership likely underestimates the long-term savings impact of a program, so alternative strategies need to be utilized to capture spillover effects and accurately report long-term savings impacts. Furthermore, clients and evaluation stakeholders need to be accepting of the partial view of evaluated net savings that this approach is capable of delivering, at least initially.
4. Evaluators should outline protocols, in advance, for providing feedback to customers when measures or systems are not operating optimally. Protocols might include factors such as: when and how the information gets relayed to customers, who communicates the information, what resources are used assist the customer in understanding the results and

potential modifications they can make, and what constitutes the final savings for evaluation reporting purposes.

Conclusions

Real-time evaluation has both benefits and drawbacks. Benefits include more reliability in customer survey responses regarding the activities they would have been likely to undertake in the absence of a program; opportunities to maximize energy saving impacts during program implementation to capture additional efficiencies; and opportunities to be flexible in program implementation to achieve energy savings goals. Drawbacks of real time evaluation include the simultaneous burden of multiple interactions with customers; the inability to capture savings derived from spillover effects; and potential for understating total actual benefits derived from program investments. Nevertheless, real-time evaluation is possible – and can be employed under reporting constraints, to more proactively ensure maximization of program benefits. Real time evaluation elements can also be integrated into complete evaluation plans, which include longer term, retrospective assessments, to counter balance some of the drawbacks of this approach.