Meeting the New York Reforming the Energy Vision Evaluation Challenge

National Grid – Joe Dolengo, Amy Dickerson & Tom Coughlin; ERS – Sue Haselhorst

The New York Reforming the Energy Vision (REV) plan has lofty and inspirational goals to transform energy efficiency programs into dynamic offerings with a deeper focus on customer engagement. Current evaluation practices for determining the reported energy and demand savings claimed from energy efficiency programs must also change, if they are to contribute to the evolving vision. This paper presents the methods and preliminary results of a retooled impact evaluation designed to provide fresh and useful feedback to program implementers and positively engage customers, as well as determine a rigorous and reliable program realization rate.

The Evaluation Challenge

The traditional evaluation paradigm is being challenged by the REV strategy of timely program feedback and deeper customer engagement. Impact evaluations, while often rigorous, have been troubled by several challenges:

1) Results are typically reported multiple years after the projects have been installed, which can render evaluation-produced program recommendations less meaningful.
    a. The processes or implementation practices that affect the evaluation results have often changed since the projects were installed.
    b. The vendors that implemented the projects and/or the technical staff or processes that were used to calculate the savings have often changed since the projects were installed.
    c. The technologies in the program have often changed since the projects were installed.
2) A realization rate alone does not tell implementation staff what needs improvement, which is compounded by changes such as those listed above.

In years when programs or technologies changed very little, these challenges had less impact. Today, however, energy efficiency program goals are increasing in many states concurrent with the advancement of building energy codes and appliance standards programs. This backdrop has required dynamic programs, where program and technology changes are the norm. This is especially true in New York State.

The Evaluation Path Less Traveled

Under REV, evaluations are directed to be "designed and implemented to yield timely information that [feeds into] the annual iterations of utility programs." National Grid's commercial and industrial (C&I) evaluation study manager commissioned ERS to conduct an innovative evaluation, measurement, and verification (EM&V) study that would revisualize an impact evaluation in ways that meets this REV challenge. This particular evaluation targeted National Grid's energy efficient C&I lighting offerings with the intention of doing a "Real Time" C&I Evaluation and incorporating additional field observations that would yield useful findings for program implementers.

National Grid completed an impact evaluation of the Large/Medium C&I lighting program several years prior that reported information several years after program implementation. During this time period, the technology had drastically changed from about 5% to over 80% LED, which diminished the value of the findings and recommendations. It was apparent that the design of the new proposed evaluation needed rethinking to ensure more useful and timely results. Several design methodologies were considered, including M&V 2.0 techniques, billing analysis, full desk review, and/or full on-site visits. Billing analysis and M&V 2.0, a variant of billing analysis, was ruled out primarily because a previous National Grid C&I billing analysis impact evaluation had mixed results with indeterminate findings. However, another key reason billing analysis was dismissed is that the sudden and wide-scale deployment of LED technology raised questions about how this technology was working in the field, which could only be answered with on-site observation and customer feedback.

This newer type of impact evaluation was built on traditional tools (i.e. spreadsheet analysis methods and data loggers) but was reimagined as follows:

- The first imperative of the design was to shrink the time interval between installation and evaluation so that the findings would reflect program implementation as it was happening. The team set a goal of getting into the field an average of 3 months after measure installation.
- A second imperative was to be efficient; the evaluation needed to add value without adding significantly to the evaluation budget. This required rethinking the sample design to allow for efficient selection of sites without the full knowledge of the population. A traditional evaluation can have a very efficient sample design because it is developed from the actual tracking data for the full period of the evaluation. The team devised an innovative rolling sampling design, which allowed for unbiased selection of sites on a quarterly basis with an efficient true-up sample in the last quarter.
- A third imperative was to bring back findings that would inform program implementers of how well the installations were performing and being perceived by customers. The team examined the tracking data and noted the remarkable transition from linear fluorescents to LEDs since the last evaluation and decided to focus on how well the technology was working in the field. In discussions with program staff, it was noted that the prescriptive versus custom processes were in a state of flux. As a result of this input, the on-site protocols were revised to include a custom versus prescriptive incentive level comparison, an on-site lighting quality audit, and a lost opportunity assessment.
- A fourth imperative was to improve communications with program implementation. If the information is gathered, but not communicated and received by implementers, the whole benefit of real-time evaluation is missed. Regular reporting (quarterly) was built into the schedule as a mechanism for fast feedback. The impact team also experimented with delivery using memos, bulletins, and in-person meetings as forums for conveying the results.

The final evaluation design was based on a mix of desktop reviews and site evaluations in order to balance time and cost constraints.

Sample Design: Freeing the Constraints of Time

In order to conduct a real-time evaluation it was very important that the sample design be constructed in a way that freed up the evaluators to perform a more timely study. The traditional method is to wait until

the program year is completed, design the evaluation months later, then start desktop reviews or site visits a couple months later. With the variable of time being the largest foe to conducting a real-time evaluation, the sample design was the first tradition that needed to be broken.

This lighting evaluation was a perfect candidate to test a "rolling sample" evaluation design that allows for quicker evaluation of installations and more timely feedback to implementation. The main difference is that sites are chosen after a quarter is completed compared to waiting for the end of a program year to choose sites to be evaluated. Figure 1 illustrates the timeline for this evaluation's rolling sample design and a traditional evaluation's timeline. Note that this was for a Lighting evaluation study and the possibilities for quick evaluations may be more limited for other seasonal based technologies.

Figure 1. Traditional vs. Rolling Sample Schedule

The first step in designing the rolling sample was to conduct a traditional stratified-ratio estimation sample design using a proxy population. The proxy population in this case was the program participants from the prior 12-month period as represented in program tracking data. The sample design targeted ±15% precision by track with a combined precision of ±10% at the 90% confidence level for the year evaluated (with no specific target by quarter). The proxy sample design established stratum boundaries and probability of selection values by stratum. At the conclusion of each quarter, the study manager gathered the tracking data for the quarter. The stratum boundaries and probability of selection values were than applied, and sites were selected for either desktop reviews or on-site M&V in that quarter. In the final quarter, the overall precision was satisfied by reviewing the population for all four quarters and adjusting the final sample selection to meet the precision targets. This allows for the flexibility of incorporating more or fewer sites than previous quarters to achieve the target precision.[1]

Desk Reviews: Telling the Inside Story

For this study, there were over 90 projects selected for desk reviews. This effort included a review of the project documentation and data from the projects' installation contractors, equipment manufacturer cut sheets, and engineering calculations. National Grid has a very thorough work and document management system as well as procedures for their energy efficiency programs, so this made it much easier to obtain

[1] The reader can refer to "It's About Time: Doing Integrated Real-Time Impact Evaluations," Sue Haselhorst and Joe Dolengo, ACEEE Summer Study 2016, for a more detailed discussion regarding the sample design.

project information. In order to have a timely evaluation, it was critical that the EM&V firm have quick turnaround of data and information by National Grid's evaluation study manager. Additionally, close collaboration was needed between the evaluation study manager and the energy efficiency sales and implementation staff in order to gather documentation quickly and to get a deeper understanding of each evaluated project in a short timeframe.

Results for this part of the evaluation were as expected, based on previous evaluations for the C&I lighting program in National Grid's NY territory. Documentation and engineering calculations showed that the calculated kW and kWh saved generally matched New York's Technical Resource Manual (NY TRM) calculations. Any given project ranged from an 80% realization rate to 95% for kW and kWh. Figure 2, below, is an illustration of the C&I Lighting formula from the NY TRM. This was used as a basis for the realization rate estimates based solely on documentation.

Figure 2. Lighting Formula

Method for Calculating Annual Energy and Peak Coincident Demand Savings

Annual Electric Energy Savings

$$\Delta kWh = \left[\frac{(W \times units)_{baseline} - (W \times units)_{ee}}{1{,}000}\right] \times hrs_{operating} \times (1 + HVAC_c)$$

Peak Coincident Demand Savings

$$\Delta kW = \left[\frac{(W \times units)_{baseline} - (W \times units)_{ee}}{1{,}000}\right] \times (1 + HVAC_d) \times CF$$

National Grid's tracking calculations for their lighting projects use site hours for calculations of kWh saved in the formula above. These hours are based on customer discussions at the time of project creation and reconfirmed at the time of installations or during post inspections. The evaluation used the tracking project site hours to compare to the NY TRM hours of use by building type. Additionally, the NY TRM has factors for HVAC interactive effects that were used for evaluation of kW and kWh tracking savings as compared to the NY TRM.

Figure 3, below, is an illustration of some of the building types listed in the NY TRM.

Figure 3. NY TRM Excerpt Operating Hours[2]

Operating Hours

The average ***lighting operating hours*** are defined by building type, as shown in the table below. These are typical average values for the building types shown. Use building specific operating hours where available.

Facility Type	Lighting Hours (hrs/yr)	HVAC Int	Facility Type	Lighting Hours (hrs/yr)	HVAC Int
Auto Related [64]	2,810	AR	Manufacturing Facility	2,857	Ind
Automotive / Transportation Service or Repair Facility (24/7)	8,760				
Bakery	2,854	FS	Medical Offices	3,748	SOfc
Banks	3,748	SOfc	Motion Picture Theatre	1,954	Asy
Church	1,955	Rel	Multi-Family (Common Areas)	7,665	MFL
College– Cafeteria(1)	2,713	FS	Museum	3,748	Asy
College – Classes/	2,586	CC	Nursing Homes	5,840	MFL

An interesting finding from this comparison was that building hours of use from tracking site hours analysis did not neatly match the NY TRM. From discussions with various implementation staff and site visits it seems that there could be high variability depending on what spaces were affected by the project. For example, conference rooms have a vastly different occupancy schedule from the overall building schedule and hours of use. Thus, just because a project is created for an office does not mean that the project's hours of use will be close to the office building type listed in the NY TRM if a larger proportion of the project is for conference rooms, bathrooms, or for other spaces with highly variable occupancy.

To increase the value of the desk reviews, this study also included an analysis of processes to improve the turnaround time for feedback to implementation staff on how well their internal procedures were being followed in order to have accurate energy saving estimates. Projects were analyzed for document completeness, and the evaluation study manager informed the implementation staff on project "completeness" metrics and gaps. Now that projects were being evaluated closer to the time of the installations it became much easier to get project details and documentation compared to traditional evaluations. It also was easier to determine the lighting project's design considerations or why certain documentation was missing and then to track down the missing information. For a traditional evaluation, if the project being evaluated is a year or more after its installation there is a much higher probability that the vendor who did the project either won't have time to go back or won't have the ability to find missing documentation. With longer time gaps, the National Grid staff involved with the particular project would not have the ability to recall project nuances, or the staff that had been involved may have been assigned to new responsibilities within the company.

Another innovative evaluation method was used in which the projects were analyzed in various ways beyond a traditional impact evaluation. The incentive design for projects was reviewed as well as its application track. In some of the projects it was found that less engineering could have been applied and more prescriptive approaches could have been used. The desk reviews helped inform National Grid's EE program designers on perhaps a more effective use of staff time or application tracks and other

[2] New York Standard Approach for Estimating Energy Savings from Energy Efficiency Programs – Residential, Multifamily, and Commercial/Industrial Measures Version 3, p 196.

technology opportunities. This effort also attempted to look at whether certain lighting measure alternatives could have been used more appropriately for the project.

Conducting the desk review first allowed for more informed site visit designs for those projects that were chosen for site reviews and data logging. Evaluators were able to design more comprehensive customer questions or were able to consider other project questions that were informed by a thorough desktop evaluation review of the project.

REV NY and Customer Engagement

Once a desk review was completed for a project that was also chosen for a site visit, the customer was contacted to schedule the on-site evaluation. This evaluation was innovative in various ways, but one of the more interesting aspects was a metric that is of high importance in REV NY – customer engagement.[3]

This evaluation study was timely and focused on REV NY principles, and as a result there was a much deeper focus on customer engagement. By having the facility site visits scheduled closer to the time of installations (and the work performed still fresh in the customer's mind) it became much easier to schedule site visits and to coordinate communications with the customer. National Grid implementation staff also benefited from the evaluations because they received more thorough and timely feedback from the customer on their experience.

The National Grid evaluation study manager worked closely with ERS on the design of the site visits to ensure that customer engagement was of prime focus. Typically in C&I lighting site evaluation visits, the evaluators will arrive at the customer's site with the main focus of asking facility personnel about the lighting usage and schedules in order to determine appropriate locations for lighting logger installations (if installing loggers) and learn hours of use and technologies installed. This study went beyond this by incorporating process evaluation into the impact evaluation and incorporating customer engagement/satisfaction observations.

Customers were asked questions regarding the lighting design and installation as well as satisfaction regarding the performance of the lighting. Additionally, customers were asked to give feedback on any recommendations for the program implementers that could help improve the design of the lighting incentive programs as well as the application process.

Lighting level satisfaction questions and feedback were compared against more quantitative lighting level approaches for the site visits. Figure 4 shows an example of the comparison of site-measured light levels to a typical lighting level standard (IESNA in this example). In general, the program-incentivized lighting fell within an acceptable range, neither under nor over lit.

[3] "Reforming the Energy Vision (REV) Working Group I: Customer Engagement; Staff Report on the Work of the Customer Engagement Committee," July 8, 2014. https://www3.dps.ny.gov/W/PSCWeb.nsf/96f0fec0b45a3c6485257688006a701a/853a068321b1d9cb85257d100067b939/$FILE/WG%201_Customer%20Engagement_Final%20Report%20&%20Attachments.pdf.

Figure 4. Average Site-Measured Light Levels Compared to IESNA Standards

Location	Task Category	Average Measured Light Level (FC)	IESNA Recommended Light Levels Range (FC)
Shed	Inactive storage	4	2–5
Back room	Inspection	64	50–100
Hall	General	27	20–50
Lounge area	Dining	5	5–10
Register	Cashier	40	20–50
Billet room mezzanine	Mechanical	3	5–10
Guestrooms – bedside	Miscellaneous	17	20–50
Guestrooms – desk	Reading	38	20–50
Breakfast room	Dining and reading	21	10–20
Manufacturing and storage	Moderate assembly	35	50–100
Patient rooms	Reading	26	20–50
Hallways	Passage	23	10–20
Dutch room	Conference	30	20–50
Kitchen	Kitchen	65	50–100
Fourth-floor hallway	Hallway	5	5–10

Lighting Lost Opportunities

As part of the REV concept of going deeper with customer engagement and the energy efficiency opportunities, the site visits also incorporated an analysis of energy efficiency opportunities that could have been proposed as part of, or in addition to, the lighting measures that were already installed.

In particular, analyses of the lighting projects were performed at the sites to determine the potential for lighting controls at the facility if none had been installed. Hours of use within the various spaces at the facility were already analyzed as part of the initial facility walk-through and remote reviews, so this was combined with an analysis of presence or lack of controls for those spaces.

Communicating with Implementation Staff

A key factor for the success of evaluations is to communicate the findings to the program administrator's implementation and program design staff in a timely fashion. Traditional evaluations are completed in a timeframe that makes it more difficult to communicate results due to the programs changing by the time evaluations are completed. So in these cases, more focus is on the realization rates rather than on program-design feedback.

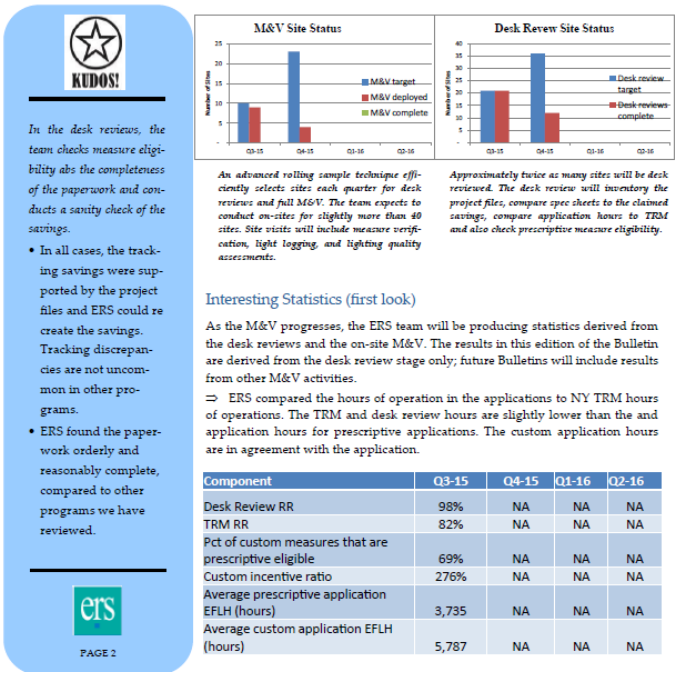
With field evaluation results available 6 months after implementation, the method for presenting information also required rethinking. In a traditional evaluation with evaluation results delivered one to several years after implementation, the results that are presented are fully formed and final and the passing of time takes the sting out of negative findings.

The evaluation team needed to consider how best to communicate results that were preliminary and subject to change (since in the first quarter only a fourth of the site visits had been completed) yet fresh

enough to indicate that action can and should be taken (which can cause distress in an organization). Some thought an experimentation was required before presenting implementers with "partial" results and findings that showed potential for improvements.

Figure 5 presents an excerpt from one of the experimental quarterly bulletins issued during the course of the study. The bulletins were intended for wide distribution and in an easy-to digest-format presenting findings as they emerged. The evaluation team found that this did not work very well because it could not, by its nature, provide a context for the results and caused concerned that the material could be misconstrued within the organization. The evaluation team found that detailed quarterly memos were useful for a core viewership, but the most effective communication was delivered in-person.

Figure 5. Excerpt from the Quarterly Bulletin



KUDOS!

In the desk reviews, the team checks measure eligibility abs the completeness of the paperwork and conducts a sanity check of the savings.

- In all cases, the tracking savings were supported by the project files and ERS could re create the savings. Tracking discrepancies are not uncommon in other programs.
- ERS found the paperwork orderly and reasonably complete, compared to other programs we have reviewed.

ers

PAGE 2

An advanced rolling sample technique efficiently selects sites each quarter for desk reviews and full M&V. The team expects to conduct on-sites for slightly more than 40 sites. Site visits will include measure verification, light logging, and lighting quality assessments.

Approximately twice as many sites will be desk reviewed. The desk review will inventory the project files, compare spec sheets to the claimed savings, compare application hours to TRM and also check prescriptive measure eligibility.

Interesting Statistics (first look)

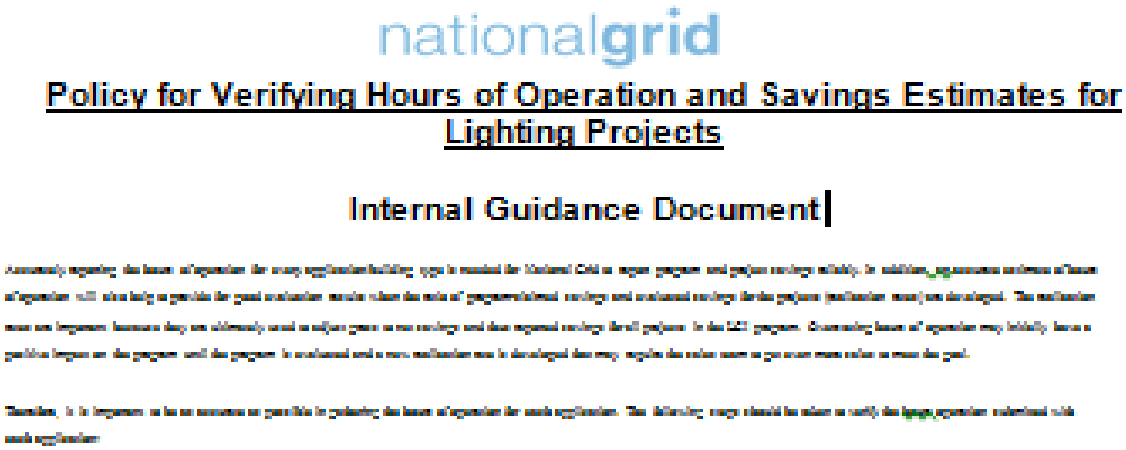
As the M&V progresses, the ERS team will be producing statistics derived from the desk reviews and the on-site M&V. The results in this edition of the Bulletin are derived from the desk review stage only; future Bulletins will include results from other M&V activities.

⇒ ERS compared the hours of operation in the applications to NY TRM hours of operations. The TRM and desk review hours are slightly lower than the and application hours for prescriptive applications. The custom application hours are in agreement with the application.

Component	Q3-15	Q4-15	Q1-16	Q2-16
Desk Review RR	98%	NA	NA	NA
TRM RR	82%	NA	NA	NA
Pct of custom measures that are prescriptive eligible	69%	NA	NA	NA
Custom incentive ratio	276%	NA	NA	NA
Average prescriptive application EFLH (hours)	3,735	NA	NA	NA
Average custom application EFLH (hours)	5,787	NA	NA	NA

At the time of this evaluation study, program implementation staff were engaging in process excellence teams with the intent of finding program process gaps. The main metrics being analyzed by these teams were lighting project lead times and overall program goal attainment. This made for an opportunity for this evaluation to provide additional metrics to these process excellence teams that were targeting accuracy of the project-claimed savings and documentation completeness. As a result of this effort, the sales staff and evaluation program managers were able to create new processes that contribute to the improvement of claimed savings accuracy. One particular example was a "Project Hours of Use Guidelines" document that was created to provide feedback regarding potential issues to watch for with facility hours of use when creating a new lighting project. See Figure 6 for an illustration of this.

Figure 6. Excerpt from Internal Guideline Document



nationalgrid

Policy for Verifying Hours of Operation and Savings Estimates for Lighting Projects

Internal Guidance Document

Conclusions

National Grid's revamping of the traditional impact evaluation set out to provide timelier, fresher, and more useful findings to the implementers as well as fulfill the regulatory requirement for rigorous impact evaluations. Important findings and actions emerged *within* the program year evaluated, including a reassessment of prescriptive versus custom measure allocations and a process excellence metric. Implementers have been briefed on preliminary findings in a workshop setting. They were reassured and happy to find that the transition to LEDs has been met with success in the field. The draft final report is underway within 6 months of the conclusion of the program year evaluated. The sample precision targets for the realization rate have been met within the original sampling parameters. Sister program administrators (PAs) in a neighboring state have taken notice and incorporated the lighting quality audit in an evaluation of an upstream lighting program, and they are including rolling samples in a new impact evaluation framework. All in all, the revamping of the impact evaluation to meet the "vision" of REV has been a success.