Clean Energy Fund Commercial Chapter Impact Evaluation

Real Time Energy Management (RTEM) Program

(2017 – Q4 2021)

Final Report

Prepared for:

New York State Energy Research and Development Authority
Albany, NY

Dana Nilsson
Senior Project Manager, NYSERDA

Prepared by:

DNV

Ben Jones, Vice President, DNV
Alain Tayoun, Senior Engineer, DNV
Paula Ham-Su, Principal Data Scientist, DNV

NYSERDA Contract 104552 September 2023

Record of Revision

RTEM Impact Evaluation Final Report
RTEM Impact Evaluation Final Report September 2023

Revision Date	Description of Changes	Revision on Page(s)
9-15-23	Original Issue	Original Issue
9-21-2023	Revised to correct MWH and MMBTU energy savings values in tables ES-1 and 2-4 as well as two associated text references. There is no impact to any other values included in the report.	3, 17

Notice

This report was prepared by DNV in the course of performing work contracted for and sponsored by the New York State Energy Research and Development Authority (hereafter "NYSERDA"). The opinions expressed in this report do not necessarily reflect those of NYSERDA or the state of New York, and reference to any specific product, service, process, or method does not constitute an implied or expressed recommendation or endorsement of it. Further, NYSERDA, the state of New York, and the contractor make no warranties or representations, expressed or implied, as to the fitness for particular purpose or merchantability of any product, apparatus, or service or the usefulness, completeness, or accuracy of any processes, methods, or other information contained, described, disclosed, or referred to in this report. NYSERDA, the state of New York, and the contractor make no representation that the use of any product, apparatus, process, method, or other information will not infringe on privately owned rights and will assume no liability for any loss, injury, or damage resulting from or occurring in connection with the use of information contained, described, disclosed, or referred to in this report.

NYSERDA makes every effort to provide accurate information about copyright owners and related matters in the reports we publish. Contractors are responsible for determining and satisfying copyright or other use restrictions regarding the content of reports that they write, in compliance with NYSERDA's policies and federal law. If you are the copyright owner and believe a NYSERDA report has not properly attributed your work to you or has used it without permission, please email print@nyserda.ny.gov.

Information contained in this document, such as web page addresses, are current at the time of publication.

Table of Contents

List of Figures

List of Tables

Executive Summary

This report describes the impact evaluation of the Real Time Energy Management (RTEM) Program that assessed the energy savings for projects installed from January 1, 2017, through December 18, 2020. All sites in the population had a minimum of one year of post-installation data at the time of the evaluation kick-off in January 2022.

RTEM[1] works in real time to monitor building systems' current and historical performance data. Data points such as set points, power loads, flow rates, temperature, and humidity, are collected and processed on-site, or on a cloud-based server, or a combination of the two to fine-tune the building energy system operations and identify capital projects. Depending on the vendor offering, RTEM systems must also provide predictive analytics, fault detection and/or diagnostics and performance optimization.

The RTEM Program provides cost-sharing incentives for both the information-gathering systems and consulting services provided by vendors reviewing the captured data. The pre-requisite consulting contract covers three years and can be extended for two additional years. Recommendations for changes are made either by vendor personnel or vendor-developed software based on the data that the system collects. Incentives are based on the project cost (30% for system installation and 30% for service years 1 through 3; reduces to 20% service cost share for years 4 and 5) and are not directly proportional to energy savings achieved by the project.

The objectives of this impact evaluation are as follows:

- Estimate the evaluated and verified gross energy impacts for RTEM projects over time, which includes electric energy (kWh) and fossil fuel energy (MMBtu) savings.
- Provide findings and recommendations to improve Program effectiveness.
- Develop the verified gross savings realization rate (SRR) or savings per unit of measure for the Program period.

Approach

The RTEM Program funded 528 sites between January 1, 2017, and December 18, 2020. These projects received incentives to help offset installation costs of either the information-gathering portion of the RTEM system, the engineering review of the data, or both. The evaluation team was tasked to conduct a review of the RTEM program to verify the gross electric and gas savings over time, develop the VGS RR,

[1] https://www.nyserda.ny.gov/ny/PutEnergyToWork/Energy-Technology-and-Solutions/Building-Operations-and-Performance/Real-Time-Energy-Management

and provide findings and recommendations to improve Program effectiveness. To date, the evaluation team has reviewed program data and performed savings analysis using a ground-up approach compiling measure details from program documents supplemented by findings from a billing analysis and from literature review of analogous evaluations. This section outlines the approach adopted in evaluating the RTEM program. Due to constraints on data access at the time of this evaluation, the evaluation primarily relied upon vendor provided service reports.

Measure-Level Approach

This method looked at sites within the population from the bottom up: Site-level savings are generated through measure-level energy savings. As part of this approach, the evaluation team conducted a detailed review of the service reports of all 528 sites. Service reports predominantly serve as a record of findings and recommendations that the vendors have made over a certain time (mostly 6 months). Several vendors, constituting over 50% of the projects in the population (310 sites), provide service reports that are less quantitative and more qualitative than the other vendors. The evaluator excluded these sites from the measure-level analysis, since additional data not contained in the vendor reports would be needed to quantify the savings impact associated with these reports. Overall, 218 sites were evaluated as part of the measure-level approach and expanded to represent the full population.

First-year savings were evaluated for each measure to study the progression of program savings over time. Installed first year savings by service year were aggregated across all measures within a project to set-up the expansion analysis.

The evaluation team also attempted to verify measure-level savings from the information provided in the reports and the trend data to compare with the vendor claimed savings. As will be discussed in depth in Section 3, the evaluation team found that there was not enough data to conduct an analysis capable of validating the claimed savings suggested in the service reports. Additional building energy data may provide a more complete analysis of savings.

Billing Analysis

The evaluation team considered using a utility data pre/post billing analysis as a primary method for reporting. This type of analysis compares energy use before and after the RTEM implementation considering the influence of weather in energy use. However, several challenges, including difficulty in obtaining granular building operational and metered energy data within the timeframe of this study, a limited number of participants with sufficient longitudinal data, and COVID-19 effects, meant that the results produced are less reliable than the measure-level approach. This billing analysis served as an additional data point to approximate the bounds of the expected savings.

Program staff provided utility information for 83 sites. The evaluation team reviewed the data and performed a quality check for completeness and usability. Upon review, 39 of the sites were removed from the evaluation sample due to a lack of the minimum required amount of data to complete the analysis, which for Year 1 was set at nine months of pre-install and nine months of post-install data. The evaluation team completed a savings analysis on 44 of the sites for Year 1, and 24 of the sites for Year 2.

For each site within the sample population, the evaluation team modeled the utility data as a function of weather data (cooling and heating degree days), and, if applicable, the post-COVID period, to establish the relationship between utility consumption and outdoor temperature.

Results

The primary focus of this study was to develop the verified gross energy savings estimates over time, savings per unit of measure and the SRR for the RTEM Program. These values were calculated by fuel type. Due to the inconsistency in reporting of heating fuels, the evaluation team opted to combine natural gas, oil, and steam into one category. Lacking a heating fuel baseline that could be extracted from the available data, the team chose a proxy for the baseline to enable an estimate of savings. Metrics calculated as a function of this estimate should not be used for reporting, they only serve to contextualize the verified gross savings. Table ES-1 outlines incremental electric and heating fuels verified gross savings over time.

Table ES-1. Verified gross savings over time

Year	Evaluated kWh Savings Percent of Baseline	Evaluated kWh Savings	Evaluated kWh/Sq.Ft	Evaluated MMBtu Savings Percent of CBECS Baseline[a]	Evaluated MMBtu Savings	Evaluated MMBtu/Sq.Ft
1	3.29%	71,520,123	0.515	0.80%	36,162	0.00026
2	1.74%	37,745,914	0.292	1.00%	45,763	0.00033
3	0.02%	353,716	0.003	0.10%	3,477	0.00002
4						
5						
Total[b]	5.05%	109,619,753	0.81	1.90%	85,403	0.00061

[a] Since heating fuel baseline energy use was not provided, the evaluation team calculated an estimated baseline based on inputs collected from the Commercial Buildings Energy Consumption Survey (CEBECS)[2].
[b] Totals may not sum due to rounding.

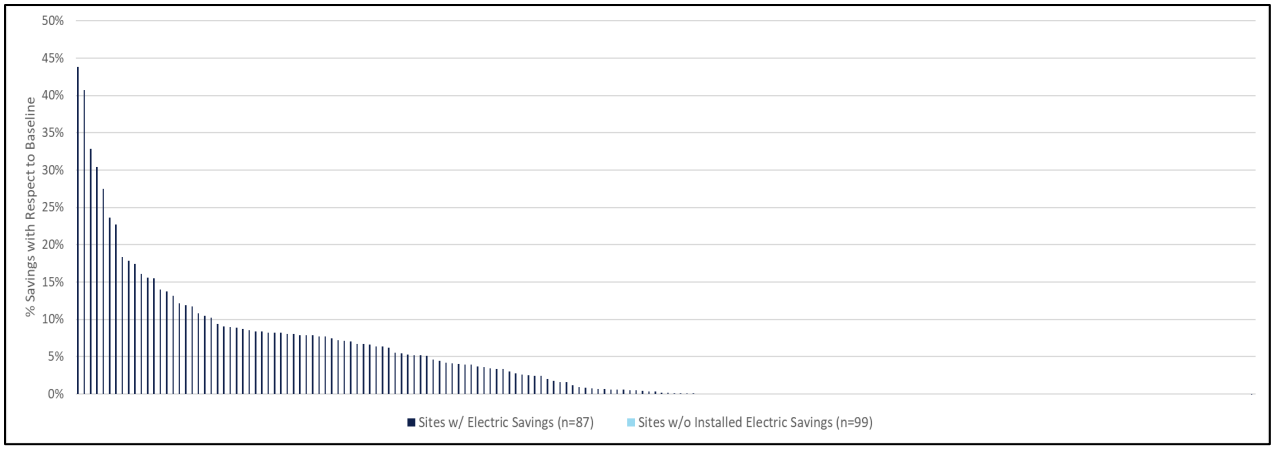
[2] 2018 CBECS consumption and expenditures. Table C25. Natural gas consumption and conditional energy intensity by census region: https://www.eia.gov/consumption/commercial/data/2018/index.php?view=consumption#c23-c32

Small sample sizes for Year 4 (n=8) and Year 5 (n=1) meant that statistically-significant determinations could not be made regarding evaluated energy savings in those years.

Overall, the Program has achieved 109,619,753 kWh of electric energy savings, or 0.81kWh/sq.ft or 5% energy reduction with respect to baseline. The program also realized 85,403 MMBTU of combined heating fuel savings or 0.61 kBTU/sq.ft. As mentioned above, since heating baseline energy use was not provided, the evaluation team estimated baseline heating fuel energy use by utilizing CBECS energy intensities by facility type. The estimated baseline was only used to approximate heating fuel savings with respect to baseline, which came out at 2%. This factor should not be used for reporting purposes and is only presented to contextualize savings. The energy savings were determined by evaluating incremental savings for each service year (savings from newly installed measures each service year) within the evaluation period.

Looking at the 218 sites reviewed, 51 sites with installed measures saved more than 5% of baseline energy. These high savings sites are balanced by 99 sites with no installed measures (recommended but not installed, therefore no energy savings) and 68 sites with savings between 0 and 5%. Figure ES-1 shows the distribution of site savings within the evaluated sample.

Figure ES-1-1. Distribution of electric savings within sample



The evaluation team calculated two separate VGS RR for this program to account for the Program's calculation update to claimed savings before and after Q4 2020. The same sample of sites was used in each estimation, with the method of program savings differing to simulate each period savings method across all sites. Table ES-2 outlines the early and late kWh and MMBtu verified gross savings realization rates.

Table ES-2. Evaluated VGS RR

Year	Sample Size	Early kWh RR[a]	Late kWh RR[b]	Early MMBtu RR[a]	Late MMBtu RR[b]
1	218	21%	40%	14%	15%
2	153	11%	21%	17%	17%
3	86	0%	0%	2%	2%
4	8				
5	1				
Total	N/A	32%	61%	33%	34%

[a] Early RR applies to projects that participated before Q4 2020.

[b] Late RR applies to projects that participated after Q4 2020.

It is important to highlight that while the evaluator utilized CBECS data to calculate the MMBTU savings as percent of baseline ratio, the evaluated saving, and therefore the MMBTU realization rates, did not. The evaluator will note that the program reported savings, in some instances, depended on CBECS metrics. The Program estimated MMBTU savings by applying a savings factor to baseline energy use which was not always available. Where that data was not available, CBECS data was used to calculate baseline energy use.

Comparison of Results with Other Literature

RTEM is relatively new to the industry, therefore, studies or evaluations that can serve as direct comparisons to its savings or longitudinal impacts are limited. In addition to the LBNL study[3], which was used as a basis for the program's claimed savings, the evaluator researched evaluations that study programs that are on the spectrum of energy management, which SEM and RTEM are both on. While SEM-type programs do not serve as a direct comparison, they do provide an idea of the savings to expect on that lower end of the spectrum (LBNL/RTEM expect 4%-9% savings whereas SEM expect 1%-5% savings). The programs reviewed follow the same basic theory that collecting and analyzing data from building systems over time will produce actionable recommendations that save energy; however, this is not a perfect comparison as the programs reviewed each differ from the RTEM program being evaluated in their implementation approach. For example, generally SEM programs tend to be less automated, less data driven and more focused on organizational recommendations than RTEM.

Furthermore, early signs show that RTEM effects are dependent on the targeted facility, with some achieving high savings and others virtually none. As RTEM and other similar offerings mature, it is

[3] https://eta-publications.lbl.gov/sites/default/files/kramer_provingbuildinganalytics_october2020.pdf

expected that Programs will get better at targeting facilities that benefit most from RTEM and potentially learn how to extract savings from ones that currently do not, ultimately achieving higher savings overall.

Lastly, the LBNL report introduces three different energy management information system (EMIS) types:

- Energy information systems (EIS)
- Fault detection and diagnostics (FDD) systems.
- Automated system optimization (ASO)

At a high level, EIS track whole building energy use to identify high-level opportunities, FDDs reduce maintenance cost, improve comfort, and find hidden energy waste, and ASO is designed to provide two-way monitoring and control of specific systems. The LBNL study focuses on EIS and FDD systems and draws on 1,333 and 509 buildings, respectively, to determine energy savings impacts after 2 years of participation. The end result is that after 2 years of service, EIS and FDD systems managed to reduce energy use by 3% and 9% respectively. ASOs were not studied due to small sample sizes. However, based on the description of ASOs, savings can be expected to fall somewhere between EIS and FDD systems.

Through conversations with program staff, the evaluation team understands that EIS as defined in the LBNL report is not eligible for the RTEM program, and that at the time of the evaluation, the RTEM vendors have two primary offerings: Software only platforms and full BMS service (software and controls). The former covers either ASO, FDD or a combination of both while the latter usually tackles both ASO and FDD. Since the Program does not currently collect information on the type of systems that are being installed, the evaluation team was unable to determine the distribution of those system types within the population and their potential overall influence on savings. As a result, no direct conclusions on expected energy savings could be drawn from the LBNL study.

Findings and Recommendations

The evaluation team understands that RTEM is no longer offered as a standalone program, so the following findings and recommendations are provided to document lessons learned and best practices that should be taken into consideration to improve the design, performance and evaluability of programs of similar nature in the future.

> **Finding 1. Program-based measurement and verification (M&V) is conducted and captures program savings. Baseline utility data has been collected for every site enrolled in the program since Q4 2020, but is only collecting post-installation data for a sample of those sites.**

Recommendation: Acquire permission and account numbers from the customer and collect two years of pre-participation utility billing data as well as key operational data, such as occupancy and operating hours at the time of enrollment where this is feasible. Ideally billing data is collected directly from the utility through electronic data interchange (EDI[xiii]) or similar approaches. Direct from utility billing data should retain meta data important for modelling savings such as read dates and whether the data for a time period is an actual or estimated read. Requesting data through EDI must be done promptly upon program enrollment, as the data will not be easily accessible later. This would also benefit future evaluations, removing the barrier of requesting permission to access utility data and securing the pre- period data from the source. Having two years of pre- and post- utility billing data allows for more accurate results during evaluation (e.g., using pooled regression analysis, the difference-in-differences method) and reduces uncertainty arising from a large percentage of estimated reads.

NYSERDA Response to Recommendation: Implemented. NYSERDA continues to take steps to calculate energy savings with reasonable and appropriate methods. NYSERDA is collecting baseline energy use for all new participants and has worked to continuously improve its savings methodologies.

Recommendation: Any analyses of energy savings by program or future evaluations should stratify by two dimensions to weight the sampled sites in order to better capture any cross-correlation of effects related to important site features. For this evaluation, the first dimension is facility type, and the second is facility size. This approach will allow for more accurate representation of the population along these dimensions. Below are the recommended stratification segments, based on the population of 528 sites that were evaluated to date. Each evaluation should assess the current population for the best stratification dimensions and segments within each dimension. Additionally, once the program increases its available data, sample and extrapolate savings within the expected fuel use type.

Facility Type:

- Commercial Office: These account for 95 out of 528 sites, and 73% of the total population energy use
- Multifamily: These account for 141 out of 528 sites, and 7% of the total population energy use.
- Other: These account for 292 out of 528 sites, and 21% of the total population energy use

Facility Size:

- Sites greater than 1,000,000 sq ft
- Sites between 100,000 sq ft and 1,000,000 sq ft
- Sites less than 100,000 sq ft

NYSERDA Response to Recommendation: Implemented. NYSERDA program implementers and evaluators will adopt this recommendation where sample size allows.

Recommendation: Collect detailed information on operational and behavioral changes from sites prior to using post-COVID-19 (2020 to present) data in billing analyses. This will allow for insight related to post-COVID operation and behavior effects as well as better differentiate use patterns and opportunities related to those employing Automated System Optimization.

NYSERDA Response to Recommendation: Pending. NYSERDA will collect this information where relevant for future evaluations.

Finding 2. In reviewing the service reports provided by the vendors, inconsistent and missing measure-level information was identified. In total, less than 50% of the sites in the population have detailed information in their reports. The evaluation found that reports are primarily generated for NYSERDA's program requirements and are often not developed with the customer in mind. In addition, NYSERDA does not currently impose any penalties on reports that show no savings or no recommendations. As a result, some vendors generate reports with the minimum accepted content to satisfy NYSERDA's requirements only. The current structure may not allow the program to get a full picture of the activities happening at sites due to the RTEM system, but the extent of this limitation is unknown. This finding is consistent with NYSERDA findings as well.

Recommendation: Simplify the format of the measure-level savings information that is collected from the vendors. Outlining what measures were recommended, their installation status, the energy savings by fuel associated with them, and a brief description or narrative of how the measure contributes to energy savings will suffice. This will allow NYSERDA to understand participant actions better, provide supporting evidence for M&V activities, and minimize the level of effort required from the vendors. NYSERDA could consider offering an incentive that scales with the savings recommended.

NYSERDA Response to Recommendation: Pending. Implementation is underway to improve customer and vendor data collection.

Finding 3. Heating fuel information is less reliable than electric information in both the service reports and billing data. Heating fuel billing data was not available for all sites with electric data. In addition, there was no indicator as to what heating fuel(s) each facility uses. For example, if natural gas data is provided for a particular site, it is not clear whether oil or steam service is also applicable to that facility. As a result, conducting a heating fuel billing analysis would have provided an incomplete picture.

The evaluation found that the program claimed oil and natural gas savings when the service reports predominantly reported natural gas savings and, in some instances, steam.

Recommendation: Similar to the electric measure and utility data recommendations above, collect natural gas billing data information as part of the program sign-on process as well. In addition, collect heating fuel measure information as part of the simplified measure collection process. This will provide greater visibility to NYSERDA on heating fuel measures and allow for improved evaluability in the future.

NYSERDA Response to Recommendation: Pending. NYSERDA is piloting use of utility data aggregators for collection of electric data directly from end customers and/or vendors. If the pilot is successful, NYSERDA will consider this for other fuels.

Finding 4. NYSERDA incentives and information continue to be transmitted through the vendor; this may perpetuate challenges obtaining energy data and associated information directly from sites especially as the program evolves to encompass RTEM and Commercial Real Estate (CRE)-Tenant.

Recommendation: Consider evaluability and evaluation approach(s) in the integration of the RTEM and CRE-Tenant programs. Obliging vendors to more thoroughly document facility utility meter(s) and corresponding RTEM monitored equipment up-front will be important for any billing analysis-based approaches. Documenting tenant space meters and linking measures in tenant spaces to RTEM affected meters will also be important, as some tenant measures may have measurable interactive effects on the whole building meter in some cases.

NYSERDA Response to Recommendation: Implemented.

Recommendation: Obtain detailed information from the vendors to better categorize the systems being implemented at each host site. These data-points include:

- Service offered: Software only/ Full Building Management System (BMS) service
- System types being implemented: Automated System Optimization (ASO) / Fault Detection and Diagnostics (FDD)/ Combination
- Systems that are being monitored and controls installed alongside/as a part of the RTEM system
- Collect metrics on these equipment that would facilitate Technical Resource Manual (TRM)-level savings calculations (Size, efficiency, age, etc.)

In the case of a full BMS service, specify what systems are connected to controls which are existing to the facility vs newly installed by vendor.

NYSERDA Response to Recommendation: Pending. NYSERDA is working to further characterize service offerings and system types across vendors.

Recommendation: Implement a system with the vendors to easily identify the most knowledgeable individual at the customer facility and collect their contact information. Having access to the appropriate contact facilitates outreach efforts that can supplement future evaluation work. The current customer relationship management (CRM) does include contact information but does not consistently include their roles within the organization. Being unable to identify the appropriate contact is generally a barrier during outreach efforts and can lead to outreach exhaustion, low response rates and incomplete information.

NYSERDA Response to Recommendation: Implemented.

Finding 5. When examining how long vendors tend to be engaged with a specific site, the trend overall showed that sites are dropping out of the program after two years.
Conversations with NYSERDA identified two main reasons for this:

1. Vendor-customer relationship ends for various reasons. This event limits long-term savings potential and reduces the persistence of operational changes made.
2. Vendor-NYSERDA relationship ends. The customer presumably continues to receive service, but NYSERDA no longer has visibility to facility improvements and therefore to the savings from measures installed in the future.

Recommendation: Where possible, document the service contract length upon entry to the program, any extensions to the contract and the reason a site stops reporting information to the program, particularly if the information exchange stops before the end of the initial three-year

period. Understanding these reasons can help in assessing the long-term impacts of the RTEM system. For example, if most stop reporting, but continue the vendor-customer relationship, persistence and long-term savings may be higher as compared to the early ending of the contract between vendors and end-users.

NYSERDA Response to Recommendation: Implemented.

Recommendation: The current evaluation found savings leveling off after two years and applies the two-year result to all sites. Supported by better information on drop out timing and reasons, future evaluations should consider whether different time frames of savings should be applied to different categories of sites based on their status with the program and possibly their reason for ending participation where applicable.

NYSERDA Response to Recommendation: Pending. NYSERDA will consider this for future evaluations that include these market actors and site types.

Finding 6. The evaluation calculated VGS RRs for RTEM as presented below. The VGS RR for electric has more than doubled since the 2021 study with electric representing a majority of program savings. The VGS RR for natural gas has remained fairly constant in the context of known data challenges. For reference, in the 2021 study, the electric VGS RR was calculated to be 20% and natural gas 42%.

Summary VGS RRs for RTEM

Time Period	Realization Rate	
	MWh	MMBtu
Q1 2017 – Q4 2020	32%	33%
Q1 2021 - present	61%	34%

1 Introduction

This report presents results of the third phase of the RTEM (Real-Time Energy Management) Program impact evaluation. Phases 1 and 2 constituted the first round of formal evaluation for the program pre-COVID period and is discussed in Section 1.3 Previous Evaluations.[4]

1.1 Program Description

RTEM works in real time to monitor building systems' current and historical performance data. Data points such as set points, power loads, flow rates, temperature, and humidity are collected and processed on-site, on a cloud-based server, or a combination of the two to fine-tune the building energy system operations and identify capital projects. Depending on the technology, RTEM systems can also provide predictive analytics, fault detection and diagnostics, and performance optimization.

EM techniques are applicable to all building types and organizational structures. Existing and new construction—including commercial, industrial, and multifamily buildings—can benefit. The initial targeted sector for the program is existing commercial buildings, with uptake likely higher in subsectors with significant existing penetration of Building Management Systems—commercial office, retail, university/college, non-profit, and healthcare. These sectors also have large, centrally managed buildings or portfolios, and therefore are more likely to have the human resources necessary to capitalize on the potential of EM.

The Program provides cost-sharing incentives for both the information-gathering systems and the vendors reviewing the captured data. Incentives are based on the project cost (30% for a three to five-year vendor contract) and are not directly proportional to energy savings achieved by the project.

1.2 Evaluation Objectives and Methods

The objectives and methods of this impact evaluation are summarized in Table 1-1.

Table 1-1. Study objectives, purpose, and methods

Objective	Purpose	Method
Evaluated gross energy impacts over time	Establish reliable savings for program participants over time. That includes consideration of measure and site lifetime savings, and measure uptake.	Collection of measure-specific information from service reports; categorization of measures; extrapolation of results to population

[4] https://www.nyserda.ny.gov/-/media/Project/Nyserda/Files/Publications/PPSER/Program-Evaluation/NYSERDA-2017-Q12020-Real-Time-Energy-Management-Impact-Final-Report.pdf

Objective	Purpose	Method
Savings per incentive and/or participant unit of measure	Calculate ratio of evaluated savings to the NYSERDA participant unit of measure (kWh saved as a percent of baseline, kWh/ft^2) to compare/contrast with other NYSERDA programs.	Utilize information from tracking data (baseline energy usage, facility area) and evaluated savings to calculate relevant metrics
Influential factors to program evaluation and findings	Determine factors that help or hinder the evaluation of the program. Identify factors that future evaluation team should take into consideration while reviewing the program.	Investigate the effects that vendor type and data availability have on the findings and the evaluation approach.

1.3 Previous Evaluations

A previous RTEM impact evaluation was completed in October 2021. The evaluation included 293 RTEM projects installed between October 31, 2016, and February 1, 2019, and contained data for the period immediately preceding COVID-19 impacts (February 28, 2020).[5] The goal of the study was to develop the SRR, verify the gross electric and gas savings, and investigate the expected life of those savings. The evaluation team adopted two separate methodologies, the first being a pre/post utility billing data analysis to estimate the verified gross savings. The second approach assessed and compared savings results to the billing analysis and built a more detailed dataset by compiling measure-level details from program documents.

The Program received utility data for 42 of the 293 participating sites to perform measurement and verification (M&V) and to estimate the Program's first year savings. The measure-level approach looked at those same 42 sites, but from the bottom up: The analysis generated site-level savings through measure-level energy savings reported in the vendor reports. This analysis was intended to provide additional context to the billing analysis results and provide additional information on the persistence and long-term savings expectations for the Program.

The evaluation team compared the vendor claimed savings to the billing analysis savings to see how the measure-level approach SRR would fare against the former approach. In addition, the evaluation team attempted to calculate measure-level savings from the information provided in the reports and the trend data to compare with the vendor claimed savings but found that there was not enough data to validate the claimed savings suggested in the service reports.

Finally, the evaluation team leveraged the collected measure-level information to examine the distribution of measures recommended and installed. This analysis separated measures into capital (new equipment)

[5] https://www.nyserda.ny.gov/-/media/Project/Nyserda/Files/Publications/PPSER/Program-Evaluation/NYSERDA-2017-Q12020-Real-Time-Energy-Management-Impact-Final-Report.pdf

and non-capital (controls, operations, and maintenance) measures and looked at TRM values for the estimated useful life of the measures. The goal of this exercise is two-pronged:

1. Identify potential savings the Program can expect beyond year one as savings from capital projects, which are more likely to be achieved in later years given a longer timeline for installation and commissioning relative to non-capital measures.
2. Identify the expected life of savings achieved by the Program.

2 Results, Findings, and Recommendations

This section presents the results, findings, and recommendations of the RTEM Program impact evaluation.

2.1 Evaluated Population

The evaluation population covers RTEM projects installed from program inception on January 1, 2017, through December 18, 2020. All sites in the population had a minimum of one year of post-installation data at the time of the evaluation kick-off in January 2022. The population includes 528 sites with 139,455,651 square feet of building area, a combined baseline energy use of 2,172,102 MWh for electric, and an estimated 4,585,749 MMBtu for heating fuels. The tracking data did not include baseline gas, oil, or steam usage; therefore, the evaluation team estimated baseline MMBtu usage by utilizing CBECS 2018[6] heating fuel energy intensity for each respective facility type. Table 2-1 summarizes RTEM participants as of January 2022 and the evaluation population.

Table 2-1. RTEM participants and evaluation population

Attrition	Application	Sites
Total in tracking data	505	966
Less than one year of data	228	438
Evaluation population	277	528

An application can contain multiple sites within it, however not all sites within an application have the same installation status. Therefore, the evaluation focused on site results rather than application level.

2.2 Results

An approach similar to the previous evaluation was proposed to achieve the study's objectives. The evaluation team planned to conduct a measure-level analysis complemented with customer interviews to get additional insight into the program's influence and collect additional site-specific information. The plan also suggested supplementing these findings with a utility data pre/post billing analysis. Due to several challenges such as low response rates and the ability to collect little qualitative value, the evaluation team decided to abandon the customer interviews and adopted the results from the measure-level approach. The findings from the billing analysis and literature from programs of comparable nature

[6] 2018 CBECS consumption and expenditures. Table C25. Natural gas consumption and conditional energy intensity by census region: https://www.eia.gov/consumption/commercial/data/2018/index.php?view=consumption#c23-c32

were used to benchmark results. The sections that follow present overall program results. Supplementary results, such as savings by vendor, facility, and measure type, can be found in Appendix B.

2.2.1 Measure-Level Approach Results

The evaluation team aimed to extract individual measure information from the service reports provided by the RTEM vendors for all 528 sites in the population. However, data extraction was only possible for a subset of sites since some vendor-specific reports did not focus on measure-level information. Table 2-2 shows the population attrition and Table 2-3 compares the population to the sample.

Table 2-2 Population attrition

Attrition	Applications	Sites
Population	277	528
Sites with no service reports/ no measure level information	142	310
Sample	135	218

Table 2-3. Evaluation population vs sample

Metric	Population	Sample
Application	277	135
Sites	528	218
Baseline annual energy use (MWh)	2.172,102	924,257
Baseline annual estimated fuel use (MMBtu)	4,585,749	1,009,414
Building area (sq.ft)	139,455,651	61,957,205
Sites with a "Software only platform" Vendor	159	74
Sites with a "Full BMS service" Vendor	341	138
Sites in earlier phase of RTEM[a]	290	124
Sites in later phase of RTEM[b]	238	94

[a] Projects that participated prior to Q4 of 2020

[b] Projects that participated after Q4 of 2020

The sum of the two vendor types does not equate to the total number of sites for both the population and sample since some sites were not classified in the tracking data. In addition, a differentiation was made between early and later sites to account for the Program's calculation update to claimed savings before and after Q4 2020, further explained in Section 3.3.3.1. The evaluation team calculated two separate VGS RR for this program

To obtain sample and population-level energy reduction, the evaluation team calculated savings from new measures installed in each year within the lifetime of a project; the savings are then aggregated across sites for each year. For example, if a project installs two measures in its first year in the program and three measures in its second year, then Year 2 savings would only show the savings from the three measures

installed in the second year. This approach was taken because of inconsistencies with how vendors report energy savings from measures implemented in previous periods. In some instances, implemented measures in the first year would still be featured and evaluated in subsequent periods, whereas in others they do not. To avoid the misinterpretation of the reporting structure with measure persistence, the evaluation calculated first-year savings for each measure and aggregated measure savings over the years to obtain total savings. Due to the inconsistency in reporting of heating fuels, the evaluation team opted to combined natural gas, oil, and steam into one category. Table 2-4 shows the population-level evaluation results.

Table 2-4. Evaluated savings

Year	Sample Size	Evaluated kWh Savings Percent of Baseline	Evaluated kWh Savings	Evaluated kWh/Sq.Ft	Evaluated MMBtu Savings Percent of CBECS Baseline[a]	Evaluated MMBtu Savings	Evaluated MMBtu/Sq.Ft
1	218	3.29%	71,520,123	0.515	0.80%	36,162	0.00026
2	153	1.74%	37,745,914	0.292	1.00%	45,763	0.00033
3	86	0.02%	353,716	0.003	0.10%	3,477	0.00002
4	8						
5	1						
Total[b]	N/A	5.05%	109,619,753	0.81	1.90%	85,403	0.00061

[a] Since heating fuel baseline energy use was not provided, the evaluation team calculated an estimated baseline based on inputs collected from the Commercial Buildings Energy Consumption Survey (CEBECS)[7].

[b] Totals may not sum due to rounding.

The sample size column indicates how many years the sites in our sample have been involved in the program. As is shown, no determinations could be made in Years 4 and 5 of the Program, since the sample size was not large enough to draw any conclusions.

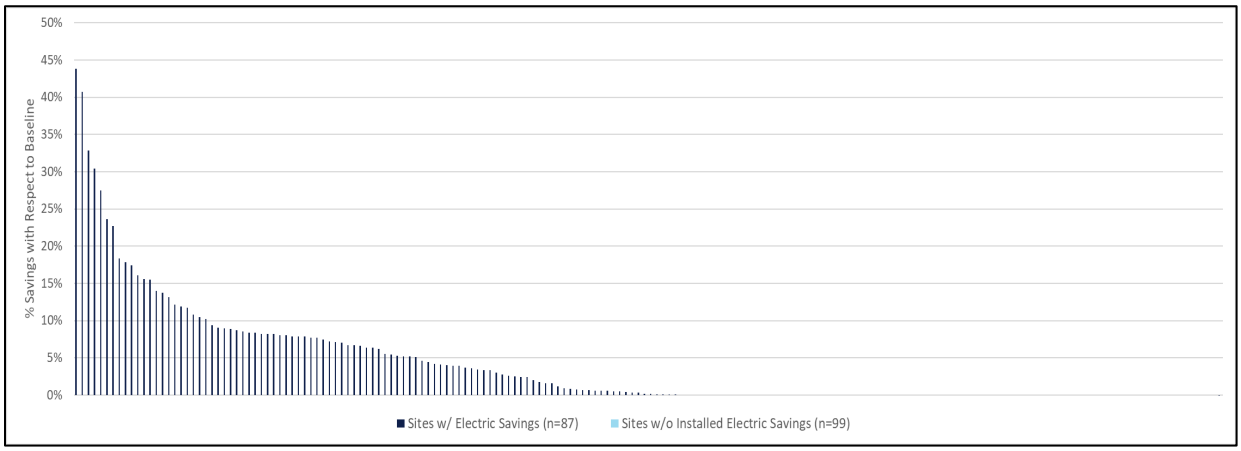
Overall, the Program has achieved 109,619,753 kWh of electric energy savings, or 0.81kWh/sq.ft or 5% energy reduction with respect to baseline. The program also realized 85,403 MMBTU of combined heating fuel savings or 0.61 kBTU/sq.ft. As mentioned above, since heating baseline energy use was not provided, the evaluation team estimated baseline heating fuel energy use by utilizing CBECS energy intensities by facility type. The estimated baseline was only used to approximate heating fuel savings with respect to baseline, which came out at 2%. This factor should not be used for reporting purposes and is only presented to contextualize savings. The energy savings were determined by evaluating incremental

[7] 2018 CBECS consumption and expenditures. Table C25. Natural gas consumption and conditional energy intensity by census region: https://www.eia.gov/consumption/commercial/data/2018/index.php?view=consumption#c23-c32

savings for each service year (savings from newly installed measures each service year) within the evaluation period.

Fifty-one sites with installed measures within the sample saved more than 5% of baseline energy. These high savings sites are balanced by 99 sites with no energy savings and 68 sites with savings between 0 and 5%. Figure 2-1. shows the distribution of site savings within the evaluated sample.

Figure 2-1. Distribution of electric savings within sample



The evaluation team calculated two separate VGS RR for this program to account for the Program's calculation update to claimed savings before and after Q4 2020. The same sample of sites was used in each estimation, with the method of program savings differing to simulate each period savings method across all sites. These two methodologies are further explained in Section 3.3.3.1. Table 2-5 outlines the early and late kWh and MMBtu verified gross savings realization rates.

Table 2-5. Evaluated VGS RR

Year	Sample Size	Early kWh RR[a]	Late kWh RR[b]	Early MMBtu RR[a]	Late MMBtu RR[b]
1	218	21%	40%	14%	15%
2	153	11%	21%	17%	17%
3	86	0%	0%	2%	2%
4	8				
5	1				
Total	N/A	32%	61%	33%	34%

[a] Early RR applies to projects that participated before Q4 2020.

[b] Late RR applies to projects that participated after Q4 2020.

It is important to highlight that while the evaluator utilized CBECS data to calculate the MMBTU savings as percent of baseline ratio, the evaluated saving, and therefore the MMBTU realization rates, did not.

The evaluator will note that the program reported savings, in some instances, depended on CBECS metrics. The Program estimated MMBTU savings by applying a savings factor to baseline energy use which was not always available. Where that data was not available, CBECS data was used to calculate baseline energy use.

Figure 2-2 and Figure 2-3 show a comparison of the evaluated site electric and heating fuel energy savings, respectively, for the sampled sites. In each of these plots, each point represents a site, with the x coordinate representing the program value for the site and the y coordinate representing the evaluated value. The line represents a slope of 1, which would be where the site would fall if the evaluated value equaled the program-reported value. Sites above the line are sites where the evaluated value is higher than the program value, and sites below the line are sites where the evaluated value is lower.

Figure 2-2. Program and evaluation electric energy savings comparison

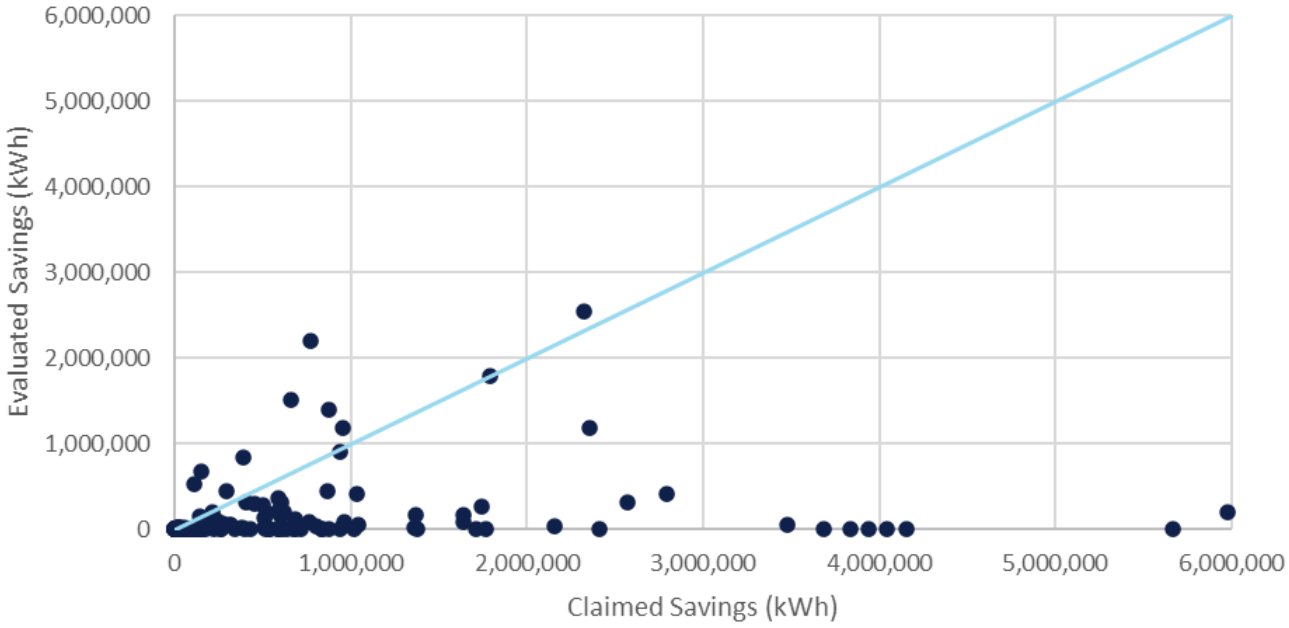


Figure 2-3. Program and evaluation heating fuel savings comparison

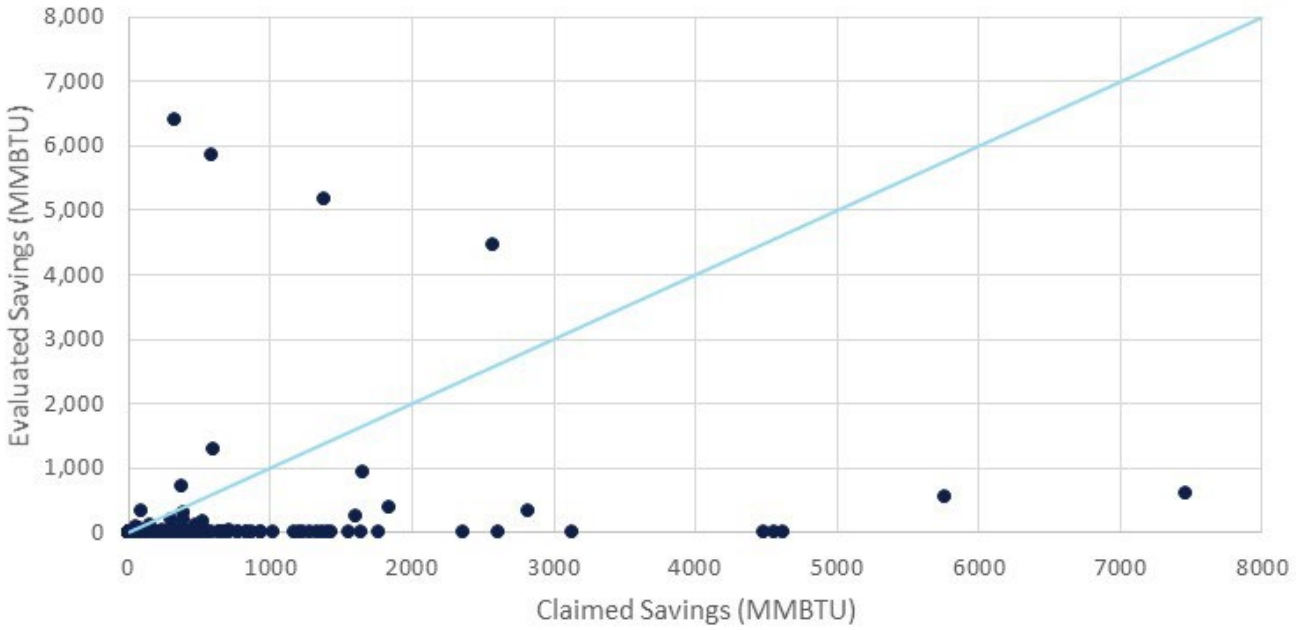


Figure 2-2 and Figure 2-3 show a high variation between the claimed and verified savings, especially for heating fuel measures. The variation is primarily caused by program estimates of savings at each site that are a simple function of square feet or baseline energy use, while the evaluation savings are based on site-specific actions taken.

2.2.2 Billing Analysis

As mentioned above, the evaluation team considered using a utility data pre/post billing analysis as a primary method for reporting. However, several challenges, including a limited amount of useable utility data and COVID-19 effects that are further discussed in Section 3.2.1, meant that the results are less reliable than the measure-level approach. The billing analysis was used to benchmark the measure-level approach. Table 2-6 shows the attrition breakdown of sites with utility billing data.

Table 2-6. Billing analysis data attrition

Category	Number of Sites
Total sites with utility data	83
Total eligible sites for Year 1 analysis (at least 9 months of pre and post data)	44
Total eligible sites for Year 2 analysis (2+ years of post-data)	24

Results from the billing analysis are shown in Table 2-7.

Table 2-7. Billing analysis results

Metric	Value

Number of sites in Billing analysis for Year 1 estimates	44
Number of sites in Billing analysis for Year 2 estimates	24
Baseline Annual Energy Use (kWh)	56,336,125
Year 1 Energy Use (kWh)	54,326,204
Year 1 Savings as Percent of Baseline	3.60%
Year 2 Energy Use (kWh)	54,223,321
Year 2 Savings as Percent of Baseline as Incremental to Year 1	0.20% (*)
Year 2 Cumulative Savings as Percent of Baseline	3.80%

(*) Statistically not significant.

The difference between the savings for Year 1 and the savings for Year 2 are statistically no different than zero. This means that the billing analysis estimates that the savings for Year 2 are effectively the same as the Savings for Year 1, and that it did not detect statistically significant incremental savings for Year 2 with respect to Year 1.

Table 2-7 shows that first-year savings for both methods are consistent. In the previous evaluation, the evaluation team was unable to verify the vendor-provided measure-level savings using system-level trend data, so this was not attempted in this evaluation. However, this comparison of the billing analysis savings with the vendor estimates suggests that, in aggregate, the vendor estimates are reasonable.

2.2.3 Comparison of Results with Other Literature

RTEM is relatively new to the industry, therefore, studies or evaluations that can serve as direct comparisons to its savings or longitudinal impacts are limited. In addition to the LBNL study[8], which was used as a basis for the program's claimed savings, The evaluator researched evaluations that study programs that are on the spectrum of energy management, which SEM and RTEM are both on. While SEM-type programs do not serve as a direct comparison, they do provide an idea of the savings to expect on that lower end of the spectrum (LBNL/RTEM expect 4%-9% savings whereas SEM expect 1%-5% savings). Table 2-8 summarizes the savings percent from these studies.

Table 2-8. Percent savings comparison with similar program

Study	% Savings
NYSERDA RTEM Evaluation (2022) – Measure-level approach (After year 3)	5%
NYSERDA RTEM Evaluation (2022) – Billing Analysis, COVID indicator var (After year 2)	4%
NYSERDA RTEM Evaluation (2020) – Billing Analysis pre-COVID (First year)	3%
Proving the Business Case for Building Analytics – LBNL (2020)[9]. (After year 2)	EIS: 4% FDD: 9%

[8] https://eta-publications.lbl.gov/sites/default/files/kramer_provingbuildinganalytics_october2020.pdf
[9] https://eta-publications.lbl.gov/sites/default/files/kramer_provingbuildinganalytics_october2020.pdf

Study	% Savings
CT EEB Strategic Energy Management (SEM) Best Practices and Evaluation[10]	3%–5%
Impact Evaluation of Commercial Strategic Energy Management - Energy Trust of Oregon[11]	1%–7%
ComEd and Nicor Gas Strategic Energy Management (SEM)[12]H	1%–1.5%
Enbridge Gas RunItRight (2017-2019)[13] [14] [15]H	2%–4%
Union Gas RunSmart Program (2018)[6] [7] [8]	4%

The programs reviewed follow the same basic theory that collecting and analyzing data from building systems over time will produce actionable recommendations that save energy; however, this is not a perfect comparison as the programs reviewed each differ from the RTEM program being evaluated in their implementation approach. For example, generally SEM programs tend to be less automated, less data driven and more focused on organizational recommendations than RTEM.

Furthermore, early signs show that RTEM effects are dependent on the targeted facility, with some achieving high savings and others virtually none. As RTEM and other similar offerings mature, it is expected that Programs will get better at targeting facilities that benefit most from RTEM and potentially learn how to extract savings from ones that currently do not, ultimately achieving higher savings overall.

Lastly, the LBNL report introduces three different energy management information system (EMIS) types:

- Energy information systems (EIS)
- Fault detection and diagnostics (FDD) systems.
- Automated system optimization (ASO)

At a high level, EIS track whole building energy use to identify high-level opportunities, FDDs reduce maintenance cost, improve comfort, and find hidden energy waste, and ASO is designed to provide two-way monitoring and control of specific systems. The LBNL study focuses on EIS and FDD systems and draws on 1,333 and 509 buildings, respectively, to determine energy savings impacts after 2 years of participation. The end result is that after 2 years of service, EIS and FDD systems managed to reduce energy use by 3% and 9% respectively. ASOs were not studied due to small sample sizes. However,

[10] https://energizect.com/sites/default/files/2021-06/C1906%20SEM%20Evaluation%20Best%20Practices%20Report_FINAL.pdf
[11] https://www.energytrust.org/wp-content/uploads/2017/03/FinalReport_EnergyTrust_CommSEM_ImpactEvaluation_wStaffResponse.pdf
[12] https://library.cee1.org/system/files/library/13219/ComEd_Nicor_SEM_EPY8_GPY5_Evaluation_Report_2016_12_16_Final.pdf
[13] https://www.oeb.ca/sites/default/files/2017-2018-DSM-Custom-Evaluation-Executive-Summary.pdf
[14] https://www.oeb.ca/sites/default/files/2018-DSM-Free-Ridership-Evaluation.pdf
[15] https://www.oeb.ca/sites/default/files/2019-Natural-Gas-Demand-Side-Management-Annual-Verification-Report.pdf

based on the description of ASOs, savings can be expected to fall somewhere between EIS and FDD systems.

Through conversations with program staff, the evaluation team understands that EIS as defined in the LBNL report is not eligible for the RTEM program, and that at the time of the evaluation, the RTEM vendors have two primary offerings: Software only platforms and full BMS service (software and controls). The former covers either ASO, FDD or a combination of both while the latter usually tackles both ASO and FDD. Since the Program does not currently collect information on the type of systems that are being installed, the evaluation team was unable to determine the distribution of those system types within the population and their potential overall influence on savings. As a result, no direct conclusions on expected energy savings could be drawn from the LBNL study.

2.3 Findings and Recommendations

The evaluation team understands that RTEM is no longer offered as a standalone program, so the following findings and recommendations are provided to document lessons learned and best practices that should be taken into consideration to improve the design, performance and evaluability of programs of similar nature in the future.

Finding 1. Program-based measurement and verification (M&V) is conducted and captures program savings. Baseline utility data has been collected for every site enrolled in the program since Q4 2020, but is only collecting post-installation data for a sample of those sites.

Recommendation: Acquire permission and account numbers from the customer and collect two years of pre-participation utility billing data as well as key operational data, such as occupancy and operating hours at the time of enrollment where this is feasible. Ideally billing data is collected directly from the utility through electronic data interchange (EDI[xiii]) or similar approaches. Direct from utility billing data should retain meta data important for modelling savings such as read dates and whether the data for a time period is an actual or estimated read. Requesting data through EDI must be done promptly upon program enrollment, as the data will not be easily accessible later. This would also benefit future evaluations, removing the barrier of requesting permission to access utility data and securing the pre- period data from the source. Having two years of pre- and post- utility billing data allows for more accurate results during evaluation (e.g., using pooled regression analysis, the difference-in-differences method) and reduces uncertainty arising from a large percentage of estimated reads.

NYSERDA Response to Recommendation: Implemented. NYSERDA continues to take steps to calculate energy savings with reasonable and appropriate methods. NYSERDA is collecting baseline energy use for all new participants and has worked to continuously improve its savings methodologies.

Recommendation: Any analyses of energy savings by program or future evaluations should stratify by two dimensions to weight the sampled sites in order to better capture any cross-correlation of effects related to important site features. For this evaluation, the first dimension is facility type, and the second is facility size. This approach will allow for more accurate representation of the population along these dimensions. Below are the recommended stratification segments, based on the population of 528 sites that were evaluated to date. Each evaluation should assess the current population for the best stratification dimensions and segments within each dimension. Additionally, once the program increases its available data, sample and extrapolate savings within the expected fuel use type.

Facility Type:

- Commercial Office: These account for 95 out of 528 sites, and 73% of the total population energy use
- Multifamily: These account for 141 out of 528 sites, and 7% of the total population energy use.
- Other: These account for 292 out of 528 sites, and 21% of the total population energy use

Facility Size:

- Sites greater than 1,000,000 sq ft
- Sites between 100,000 sq ft and 1,000,000 sq ft
- Sites less than 100,000 sq ft

NYSERDA Response to Recommendation: Implemented. NYSERDA program implementers and evaluators will adopt this recommendation where sample size allows.

Recommendation: Collect detailed information on operational and behavioral changes from sites prior to using post-COVID-19 (2020 to present) data in billing analyses. This will allow for insight related to post-COVID operation and behavior effects as well as better differentiate use patterns and opportunities related to those employing Automated System Optimization.

NYSERDA Response to Recommendation: Pending. NYSERDA will collect this information where relevant for future evaluations.

Finding 2. In reviewing the service reports provided by the vendors, inconsistent and missing measure-level information was identified. In total, less than 50% of the sites in the population have detailed information in their reports. The evaluation found that reports are primarily generated for NYSERDA's program requirements and are often not developed with the customer in mind. In addition, NYSERDA does not currently impose any penalties on reports that show no savings or no recommendations. As a result, some vendors generate reports with the minimum accepted content to satisfy NYSERDA's requirements only. The current structure may not allow the program to get a full picture of the activities happening at sites due to the RTEM system, but the extent of this limitation is unknown. This finding is consistent with NYSERDA findings as well.

Recommendation: Simplify the format of the measure-level savings information that is collected from the vendors. Outlining what measures were recommended, their installation status, the energy savings by fuel associated with them, and a brief description or narrative of how the measure contributes to energy savings will suffice. This will allow NYSERDA to understand participant actions better, provide supporting evidence for M&V activities, and minimize the level of effort required from the vendors. NYSERDA could consider offering an incentive that scales with the savings recommended.

NYSERDA Response to Recommendation: Pending. Implementation is underway to improve customer and vendor data collection.

Finding 3. Heating fuel information is less reliable than electric information in both the service reports and billing data. Heating fuel billing data was not available for all sites with electric data. In addition, there was no indicator as to what heating fuel(s) each facility uses. For example, if natural gas data is provided for a particular site, it is not clear whether oil or steam service is also applicable to that facility. As a result, conducting a heating fuel billing analysis would have provided an incomplete picture.

The evaluation found that the program claimed oil and natural gas savings when the service reports predominantly reported natural gas savings and, in some instances, steam.

Recommendation: Similar to the electric measure and utility data recommendations above, collect natural gas billing data information as part of the program sign-on process as well. In addition, collect heating fuel measure information as part of the simplified measure collection

process. This will provide greater visibility to NYSERDA on heating fuel measures and allow for improved evaluability in the future.

NYSERDA Response to Recommendation: Pending. NYSERDA is piloting use of utility data aggregators for collection of electric data directly from end customers and/or vendors. If the pilot is successful, NYSERDA will consider this for other fuels.

Finding 4. NYSERDA incentives and information continue to be transmitted through the vendor; this may perpetuate challenges obtaining energy data and associated information directly from sites especially as the program evolves to encompass RTEM and Commercial Real Estate (CRE)-Tenant.

Recommendation: Consider evaluability and evaluation approach(s) in the integration of the RTEM and CRE-Tenant programs. Obliging vendors to more thoroughly document facility utility meter(s) and corresponding RTEM monitored equipment up-front will be important for any billing analysis-based approaches. Documenting tenant space meters and linking measures in tenant spaces to RTEM affected meters will also be important, as some tenant measures may have measurable interactive effects on the whole building meter in some cases.

NYSERDA Response to Recommendation: Implemented.

Recommendation: Obtain detailed information from the vendors to better categorize the systems being implemented at each host site. These data-points include:

- Service offered: Software only/ Full Building Management System (BMS) service
- System types being implemented: Automated System Optimization (ASO) / Fault Detection and Diagnostics (FDD)/ Combination
- Systems that are being monitored and controls installed alongside/as a part of the RTEM system
- Collect metrics on these equipment that would facilitate Technical Resource Manual (TRM)-level savings calculations (Size, efficiency, age, etc.)

In the case of a full BMS service, specify what systems are connected to controls which are existing to the facility vs newly installed by vendor.

NYSERDA Response to Recommendation: Pending. NYSERDA is working to further characterize service offerings and system types across vendors.

Recommendation: Implement a system with the vendors to easily identify the most knowledgeable individual at the customer facility and collect their contact information. Having access to the appropriate contact facilitates outreach efforts that can supplement future evaluation work. The current customer relationship management (CRM) does include contact information but does not consistently include their roles within the organization. Being unable to identify the appropriate contact is generally a barrier during outreach efforts and can lead to outreach exhaustion, low response rates and incomplete information.

NYSERDA Response to Recommendation: Implemented.

Finding 5. When examining how long vendors tend to be engaged with a specific site, the trend overall showed that sites are dropping out of the program after two years. Conversations with NYSERDA identified two main reasons for this:

2. Vendor-customer relationship ends for various reasons. This event limits long-term savings potential and reduces the persistence of operational changes made.
3. Vendor-NYSERDA relationship ends. The customer presumably continues to receive service, but NYSERDA no longer has visibility to facility improvements and therefore to the savings from measures installed in the future.

Recommendation: Where possible, document the service contract length upon entry to the program, any extensions to the contract and the reason a site stops reporting information to the program, particularly if the information exchange stops before the end of the initial three-year period. Understanding these reasons can help in assessing the long-term impacts of the RTEM system. For example, if most stop reporting, but continue the vendor-customer relationship, persistence and long-term savings may be higher as compared to the early ending of the contract between vendors and end-users.

NYSERDA Response to Recommendation: Implemented.

Recommendation: The current evaluation found savings leveling off after two years and applies the two-year result to all sites. Supported by better information on drop out timing and reasons, future evaluations should consider whether different time frames of savings should be applied to different categories of sites based on their status with the program and possibly their reason for ending participation where applicable.

NYSERDA Response to Recommendation: Pending. NYSERDA will consider this for future evaluations that include these market actors and site types.

Finding 6. The evaluation calculated VGS RRs for RTEM as presented below. The VGS RR for electric has more than doubled since the 2021 study with electric representing a majority of program savings. The VGS RR for natural gas has remained fairly constant in the context of known data challenges. For reference, in the 2021 study, the electric VGS RR was calculated to be 20% and natural gas 42%.

Summary VGS RRs for RTEM

Time Period	Realization Rate	
	MWh	MMBtu
Q1 2017 – Q4 2020	32%	33%
Q1 2021 - present	61%	34%

3 Methods

The evaluation team's initial approach leaned heavily on synthesizing service report information and building a measure-level dataset while supplementing it with information collected through customer phone interviews. The former exercise would be further supplemented with a utility data billing analysis where possible. However, through conversation with Program staff, the evaluation team learned of the challenges and significant response rate issues that the Program and market evaluation faced. As a result, the evaluation team decided to forgo customer outreach and data collection to focus time and resources on a census attempt of data collection.

Section 3 describes the methods used to develop impact estimates for the RTEM population from project years 2017 through Q4 2020.

3.1 Measure-Level Approach

The evaluation team received project files supporting the 528 sampled sites. Project files included the vendor's service reports, RTEM meter trend data, project documentation, and, in some instances, calculation spreadsheets. The following section describes the steps taken by the evaluation team in conducting the measure-level approach.

3.1.1 Service Report Review

The evaluation team conducted a detailed review of the service reports of all 528 sites. The goal of this review was to understand the scope of the projects and to extract information pertaining to the recommended measures resulting from the RTEM system observations. The extracted information included:

- Recommended measure description and type (capital, controls, and O&M)
- Measure installation status and installation date
- Energy and cost savings resulting from the measure (electric, gas, oil, or steam)
- Facility and vendor type

This information was collected for all service reports within a project. Service reports predominantly cover a period of six months, with a few exceptions covering quarters. The number of service reports provided depends on how long the RTEM system has been installed. The evaluation team encountered one to seven service reports per site, covering three to five years of service. Measures were then grouped into buckets based on the service year they were installed in to determine their first-year savings. Since vendors were not consistent in their reporting, the evaluation team opted to evaluate the first-year savings for each measure to study the progression of program savings over time. For example, if a project installs two measures in its first year in the Program and three measures in its second year, then Year 2 savings would only show the savings from the three measures

installed in the second year. Please note that a calendar year is not related to a service year. A site-specific service year is the 12-month increment period that follows the installation date. Service year was used as the reference timeframe to assess the overall program savings over time. Utilizing calendar years would provide a distorted image, since the first year of a project lifecycle uncovers different measure types than subsequent years. Early recommendations tend to tackle "low-hanging fruit," whereas later ones tend to focus on larger-scale efforts.

Ultimately, installed first-year savings by service year were aggregated across all measures within a project to set up the expansion analysis.

Through the review, the evaluation team identified several vendors, constituting over 50% of the sites in the population, that provide service reports that are less quantitative and more qualitative than the other vendors. The evaluation team excluded these sites from the measure-level analysis, since no data conducive to an analysis could be extracted from the reports. Savings from these sites were captured as part of the expansion analysis. Overall, 218 sites were evaluated as part of the measure-level approach.

3.1.2 Trend Data Analysis

Following the measure-level data collection, the evaluation team proceeded to verify the energy savings reported by the vendors by conducting an analysis on the trend data provided. The following steps were taken in attempting to conduct the analysis:

- Examine the trend data to identify points that are relevant to the installed measures.
- Identify key information related to the affected mechanical equipment (size, efficiency, flow rate, etc.)
- Obtain weather data from the closest weather station for the same time as the data on hand.
- Check the measure implementation date against the time series start and end date to determine whether enough pre- and post-installation data is provided.
- Conduct analysis based on the measure description to identify whether the recommended changes are reflected in the data.

The approach described was attempted on a sub-set of sites in the sample. The analysis proved to be inconclusive for various site-dependent factors. The leading factors were sites having either pre- or post-installation data, no relevant data points to the recommended measure, missing equipment specs, and, in some instances, no trend data.

3.1.3 Method Limitations

The measure-level approach offers upsides and greater resolution into the program than a billing analysis method since it involves a deeper review of site-specific documents. This results in categorizing measures and

identifying how the savings are distributed within those measure type. On the other hand, the method also presents limitations and potential biases, which are described below:

- The evaluation team was unable to independently verify the savings presented in the service reports, as various inputs, from trend data to equipment specific information, were missing.
- The evaluation team planned to collect site-specific information pertaining to the analysis as well as measure installation status through customer interviews. However, throughout 2021 planning, the evaluation team learned of the challenges and significant response rate issues that both the Program and market evaluation faced. As a result, the evaluation team decided to forgo customer outreach, as it the process would prove challenging, yield low participation, and any information collected would likely provide little statistically significant quantifiable value needed for an impact evaluation.
- The approach relies on vendor-provided information, which presents multiple sources of potential bias:
    - The approach is limited to vendors/sites that provide measure-level savings information. Reports that did not include measure-specific details were not excluded from the sample due to lack of information. Therefore, the sample exclusively includes sites that reported measure savings. Potentially, vendors who report measures savings to NYSERDA are also providing higher-quality information to their customers than those who do not. This would create a potential for overestimating savings by relying on the vendor reports.
    - In some instances, vendors stopped providing reports to NYSERDA. In that event, the status of the relationship between the vendor and customer is unknown, and thus neither the long-term effect of the measures nor future recommendations and installations can be captured by evaluation. The implicit bias is that by using vendor reports for later years to estimate savings for the full population of sites, we overestimate savings, since some sites are not using the RTEM system for or beyond the initial three year contract.
    - The vendors are not consistent in how savings are reported. The evaluation has noted several instances where a measure that was recommended in one period no longer features in subsequent periods. The opposite is also true, where a measure is continuously reported on beyond the period where it was first recommended and installed. The reporting format was not consistent, neither across nor within the same vendor. This phenomenon impacts the evaluation of a measure's persistence since it is not clear whether the absence of a measure in a report indicates vendor reporting preference, or if the measure is no longer achieving savings. This uncertainty does not have a clear directional bias.
- Non-routine events and COVID-19 occupancy change effects are not always easy to see or control for when looking at a service report. Few service reports addressed the issue, while the majority do not mention the matter altogether. As a result, savings from an occupancy or reset measure, for example, can be greater than what is typically expected because of increased vacancies in commercial office spaces. In the early COVID years, this may have a bias for overestimation of savings, while later COVID and post-COVID years may have a bias for underestimating savings. COVID may also affect the persistence of operational changes to buildings such as set points and schedules. As commercial spaces fill up, schedules may shift to longer hours in reaction to the comfort of tenants.

3.2 Billing Analysis

The evaluation team used electric billing data collected as part of the Program's M&V process to complete a pre/post billing analysis. Program staff provided utility information for 83 sites. The evaluation team reviewed the data and performed a quality check for completeness and usability. Upon review, 39 of the sites were dropped from the evaluation sample due to a lack of the minimum required amount of data to complete the analysis which, for Year 1, was set at nine months of pre-install and nine months of post-install data. The evaluation team completed a savings analysis on 44 of the sites for Year 1 and 24 of the sites for Year 2.

For each site within the sample population, the evaluation team modeled the utility data as a function of weather data (cooling and heating degree days), and, if applicable, the post-COVID period, to establish the relationship between utility consumption and outdoor temperature.

Each customer reacts differently to outdoor temperatures. The evaluation team tested each customer's response to combinations of wet and dry bulb for cooling degree days base 60, 65, and 70, and heating degree days base 55, 60, and 65 (18 combinations total) in order to select the best model fit for each site.

The "best model" is that with the highest R-squared (R2). In these regressions, R2 is a measure of the energy consumption's response to outdoor temperatures. A customer with a high R2 has energy use that responds to weather. A site with a low R2 has energy use that does not respond to weather. In this sense, a low R2 is a valid finding.

Once the relationship between the site's energy use and outside temperature was established using actual weather data, these regressions were then used to calculate expected annual use by applying recent 10-year average weather data to the models.

```
kWh = intercept
   + Pre_Post
   + COVID
   + &Bulb._BillPeriod_CDD&CDD1
   + &Bulb._BillPeriod_HDD&HDD1
   + Year2
   + Error
```

Where:

Pre_Post

0: 6 months after measure installation ("learning period") and all months prior

1: all months after learning period

COVID

0: the months up to and including February of 2020

1: March 2020 and all subsequent months

&Bulb

Whether the best model utilized Wet Bulb or Dry Bulb outdoor temperatures

&CDD1

The cooling degree days base utilized in the best model. The CDD bases tested included 60, 65, and 70.

&HDD1

The heating degree days base utilized in the best model. The HDD bases tested included 55, 60, and 65.

Year2

0: the billing period is not in Year 2

1: the billing period is in Year 2

Error

the regression residual, or "error".

Savings were then calculated as the difference in the weather-normalized pre-installation (base case) consumption and the weather-normalized post-installation (post-case) consumption.

The models and weather-normalized consumption are illustrated in Figure 3-1 and Figure 3-2. These figures illustrate the kWh from billing data, the modeled kWh, and the weather-normal kWh that is used to estimate savings.

Figure 3-1 shows a project in the sample with the highest response to outdoor temperature. The graph shows that kWh and modeled kWh are very close to each other.

Figure 3-1. Sample customer with highest response to outdoor temperature

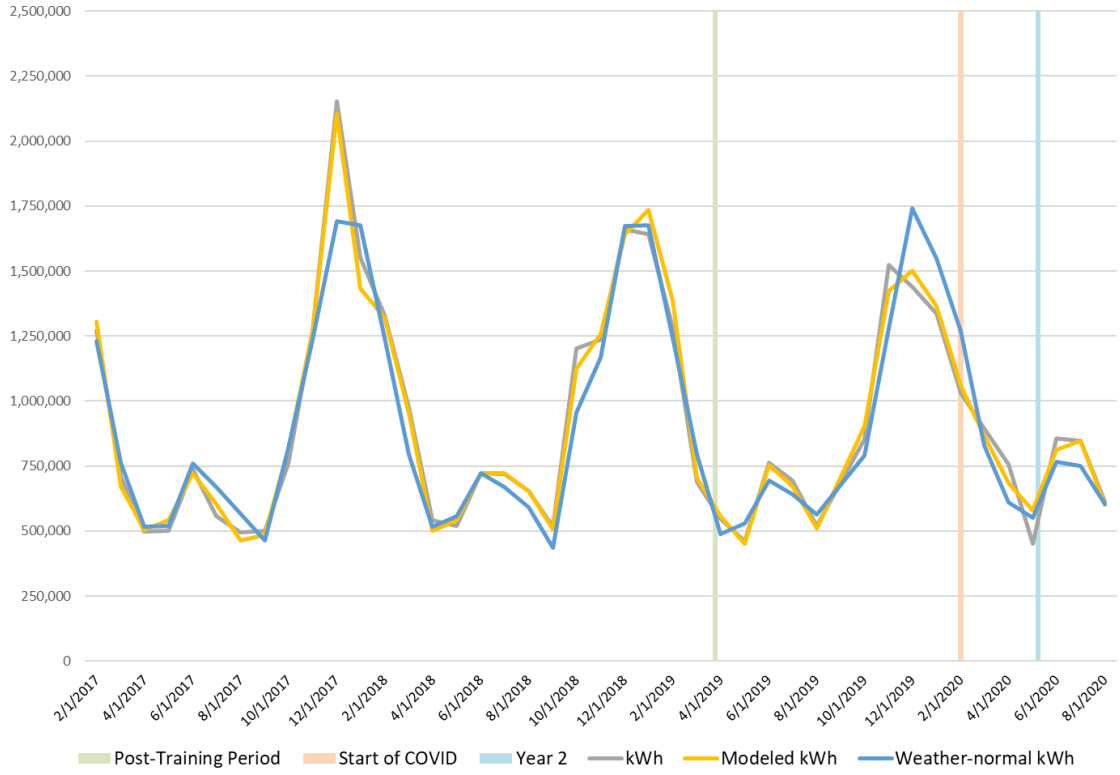
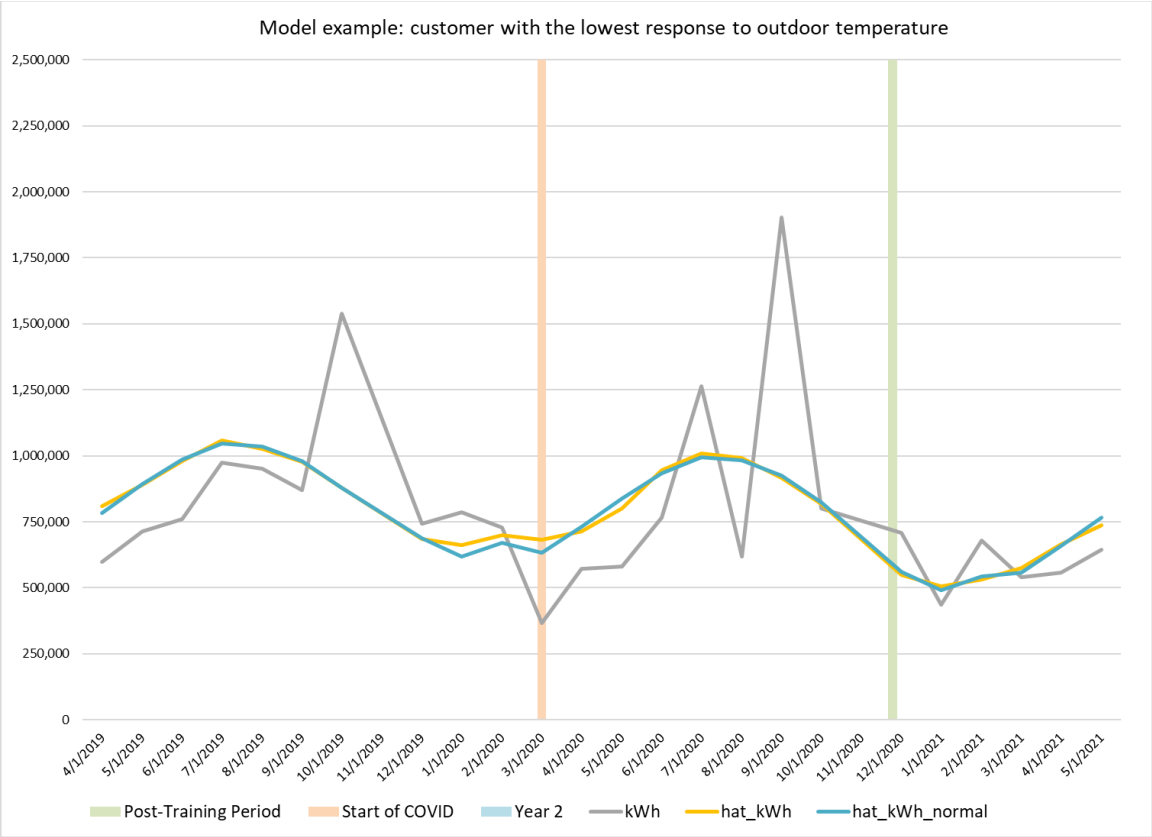


Figure 3-2 displays the kWh, modeled kWh, and weather-normal kWh for the sample customer with the lowest response to weather. The spikes and dips in energy use indicate high variability that is independent of outdoor temperature. The modeled kWh is an average use that loosely follows actual energy use.

Figure 3-2. Sample customer with lowest response to outdoor temperature



3.2.1 Method Limitations

Billing analysis and weather normalization are proven methods to estimate savings. For this study, their reliability is limited by:

- **Small sample size:** The number of sites that could be utilized in this sample was limited by the number of customers that the Program collected baseline energy for, and the number of billing periods available for the analysis, which were different for each customer. The Program has adopted a process where baseline energy use is collected for all new participants as of Q4 2020; however, only a sample of these participants are providing post-installation utility data. Both of these factors combined limit the number of participants eligible for billing analysis.
- **Biased sample:** An ideal set of customers for an evaluation study is either a census (all customers that participated in the program) or a sample selected to be representative of the program. This sample consists of customers that submitted both baseline and post-installation utility data, making it a sample that is likely biased. Biases are inherently almost impossible to measure.
- **COVID-19 impacts:** While the pandemic was incorporated into the model, we do not have enough data to fully assess whether the effect of the pandemic was fully captured in the model. Nine months of post-pandemic energy use would be ideal. The post-pandemic state is also fluid, with few commercial buildings having a consistent occupancy.

Figure 3-3. Baseline, Year 1, and Year 2 utility data availability with COVID-19 impact

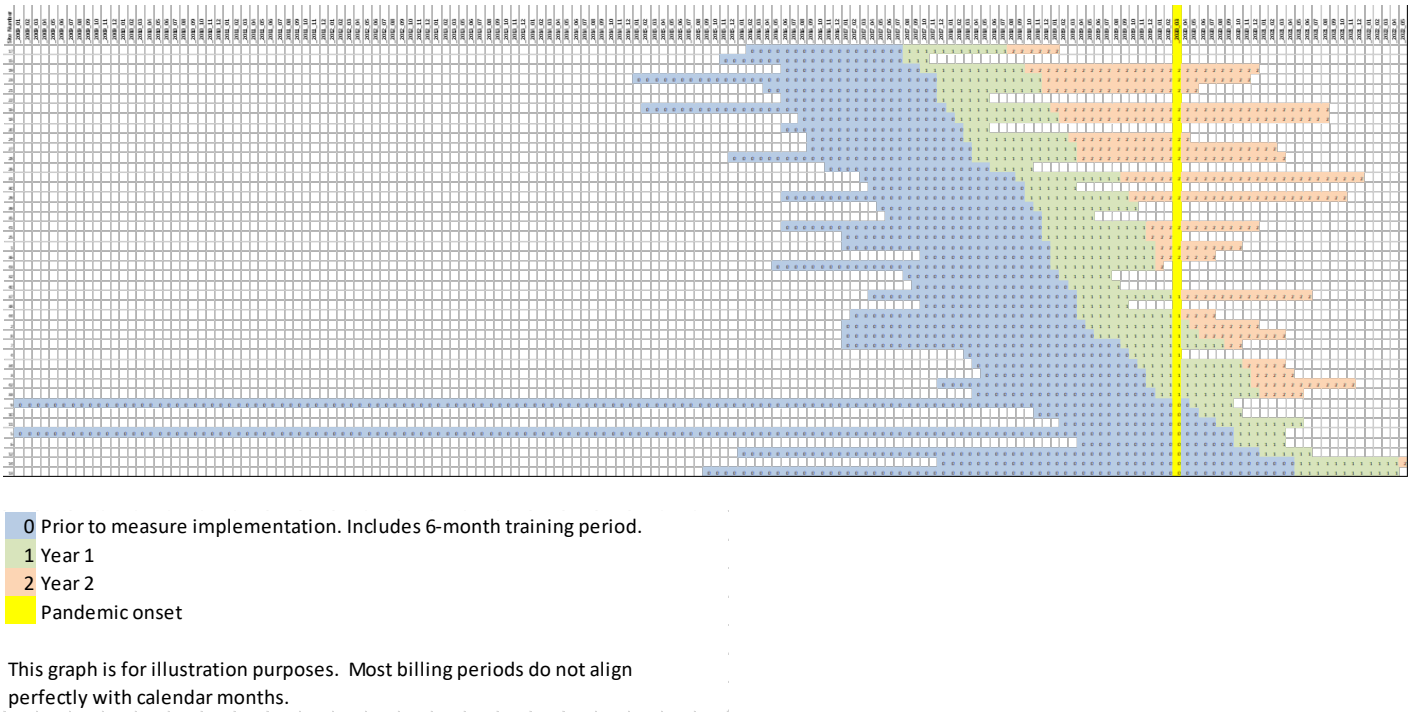


Figure 3-3 shows data availability for all 83 sites and gives an overview of how many sites have sufficient data to qualify for billing analysis. The figure also shows when we expect occupancy and therefore energy changes due to COVID-19 impacts.

3.3 Aggregate Analysis

The following describes the key steps and factors for the aggregate analysis.

3.3.1 Site Baseline Energy Consumption

For this analysis, the evaluation team used the Program-reported baseline energy use. Heating fuel baseline was not provided as part of the tracking dataset, and therefore was not included in the aggregate analysis. The evaluation team obtained heating fuel energy intensity from CBECS by facility type and applied these factors to the population to estimate baseline usage. The estimated baseline was only used to approximate heating fuel savings with respect to baseline.

3.3.2 Savings Extrapolation

The baseline and post-installation energy usage, as well as the energy savings, were established for both electric and gas usage through the measure-level approach as described in Section 3.1. As a result, savings from all 218 sites in our sample are calculated. To extrapolate to the entire population, the evaluation team assigned each of the 218 sites a sample weight based on their customer segment and size. The tracking dataset is considered as

the reference in the size weighting. Since the program had previously calculated the savings by applying a factor to the baseline energy usage, which in turn is based on square footage and the corresponding CBECS energy intensity value, the square footage was used as the proxy for a site's size as follows:

- **Small:** sites with square footage less than or equal to 100,000 sq. ft.
- **Medium:** sites with square footage between 100,000 sq. ft. and 1,000,000 sq. ft.
- **Large:** sites with square footage greater than or equal to 1,000,000 sq. ft.

The evaluation team found that two sector types are mostly represented in the sample and population: multifamily and commercial office. To properly represent the remaining sector types and avoid having the main two segments dominate the sample in weight, the remainder of the sites were grouped under "other." Table 3-1 shows the population spread across the four sector types and facility size.

Table 3-1. Facility type distribution

Evaluation Assigned Sector	Number of Facilities in Sample	Number of Facilities in Population	Facility Type Percent of Baseline within Sample	Facility Type Percent of Baseline within Population
Commercial Office	51	95	76%	73%
Other	97	292	12%	21%
Multifamily	70	141	12%	7%
Total	218	528	100%	100%

The "Other" category includes the following facility types: College/University, Commercial Retail, Food/Beverage, Government, Healthcare, Hospitality, K-12 Schools, Manufacturing, Not for Profit, and others.

Following the segmentation, the sampled sites were assigned a weight based on how many sites of the same size and sector they represent in the population. The weight is the ratio of the number of sites within a particular category in the sample to the number of sites in the same category in the population. When there are no large sites of a particular sector in our sample, the medium and large sites are grouped together. Table 3-2 shows the weight calculations.

Table 3-2. Weight calculation

Sector	Size	Number of sites in sample	Number of sites in population	Weight
Commercial Office	Small	4	1	4.0
	Medium	51	27	1.9
	Large	29	12	2.4
Multifamily	Small	62	43	1.4
	Medium	55	31	1.8
	Large	4	3	1.3

Sector	Size	Number of sites in sample	Number of sites in population	Weight
Other	Small	223	50	4.5
	Medium	68	19	3.6

Subsequently, the weight of each site was applied to both the evaluated savings and the program-level savings within the sample. The realization rate was then calculated as the ratio of the weighted evaluated savings to the weighted program savings. The evaluated program-level savings were obtained by applying the realization rates for each fuel type to the program reported savings.

3.3.3 Realization Rate

3.3.3.1 *Program Claimed Savings Calculations*

The claimed savings calculations changed over the course of the program. Therefore, to appropriately assess the performance of the program, the evaluation team calculated an "early" and a "late" realization rate that should be applied to the appropriate set of Program participants. The early realization rate applies to participants that joined the program prior to Q4 of 2020 and the late realization rate applies to participants that joined after Q4 of 2020. Table 3-3 describes the two methodologies the Program adopted to calculate claimed savings.

Table 3-3. Program claimed savings calculations

Prior to Q4 2020	After Q4 2020
Program-Reported MWh Savings	
• Multifamily: 0.1548*Baseline Energy Use • Commercial < 25,000 sq.ft: 0.096*Baseline Energy Use • Commercial > 25,000 sq.ft: 0.159 *Baseline Energy Use	• Multifamily: 0.087*Baseline Energy Use • C&I: 0.082*Baseline Energy Use
Program-Reported MMBtu Savings	
• Multifamily: 0.1548*Baseline Energy Use • Commercial < 25,000 sq.ft: 0.096*Baseline Energy Use • Commercial > 25,000 sq.ft: 0.159 *Baseline Energy Use • Natural Gas to Oil split: MF 70/30, 100/0 Commercial	Program Reported MMBtu Savings - NG: • Multifamily: 0.138*Baseline Energy Use • C&I: 0.047*Baseline Energy Use Program Reported MMBtu Savings - #2 Fuel Oil: • Multifamily: 0.138*Baseline Gallon Oil Use*0.141 • C&I: 0.047*Baseline Gallon Oil Use*0.141

The Program modified the energy savings factors applied and eliminated the size segmentation previously used. The Program also now claims oil savings for non-multifamily projects, which was not the case previously.

3.3.3.2 *Verified Gross Savings Realization Rate Calculations*

Once the evaluation team calculated savings for each of the sites within the sample, they then quantified the overall SSRs by fuel type and facility type. The facility size was embedded in the weights assigned to sites of different sizes within a particular facility type. As noted above, this breakout was chosen to properly stratify the population and be consistent with future evaluations of the Program. The RR calculation is as follows.

Notation: The following terms are used in calculating the realization rate for each fuel type and time period:

T_j = Tracking estimate of gross savings for site j

V_j = Verified estimate of gross savings for site j

W_j = Sample weight for site j

S = Number of sites in the sample

The realization rates are calculated directly:

$$RR = \frac{\sum_{j=1}^{S} W_j * V_j}{\sum_{j=1}^{S} W_j * T_j}$$

3.3.4 Precision and Limitations

Relative precision was calculated using the procedures described in Chapter 13 of the California Evaluation Framework.[16]

The evaluation was limited by the following factor:

- **Non-random sample:** The evaluation team selected the sites in the sample based on availability of data (vendor reports with measure savings). A census was attempted, however, since a number of sites did not have sufficient data required for full evaluation, it should be noted that attrition led to sites that were not selected at random within each stratum, which could result in potential bias.

[16] http://www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf

Appendix A: Glossary of Terms

census – All individuals in a group. In evaluations of energy efficiency programs, census typically refers to all the projects in a stratum of program projects.

evaluated gross savings – The change in energy consumption and/or demand that results directly from program-related actions taken by participants in an efficiency program, regardless of why they participated, as calculated by the program evaluation team.

evaluated net savings – The total change in load that is attributable to an energy efficiency program, as calculated by the program evaluation team. This change in load may include, implicitly or explicitly, the effects of free drivers, free riders, energy efficiency standards, changes in the level of energy service, and other causes of changes in energy consumption or demand.

net savings – The total change in load that is attributable to an energy efficiency program. This change in load may include, implicitly or explicitly, the effects of spillover (SO), free riders, energy efficiency standards, changes in the level of energy service, and other causes of changes in energy consumption or demand.

nonparticipants/nonparticipating – Any customer or contractor who is eligible but did not participate in the program under consideration. Nonparticipating contractors can include contractors who have never participated in the program and contractors who formerly participated prior to the year(s) being evaluated but have not participated since.

normal replacement – The replacement of equipment that has reached or passed the end of its measure-prescribed expected useful life (EUL).

overlap (OL) – The proportion of installed measures for which customers received funding from other NYSERDA programs or other sources.

participant – An end user who receives an assessment or a service provider—assessment provider, expeditor, or finance partner—associated with the program.

relative precision – Reflects the variation due to sampling as compared to the magnitude of the mean of the variable being estimated. It is a normalized expression of a sample's standard error from its mean. It represents only sampling precision, which is one of the contributors to reliability and rigor and should be used solely in the context of sampling precision when discussing evaluation results.

Appendix B: Supplementary Results

Savings by vendor, facility, and measure type are presented below.

Savings by vendor type

Figure B- 1 Incremental electric savings over time by vendor type

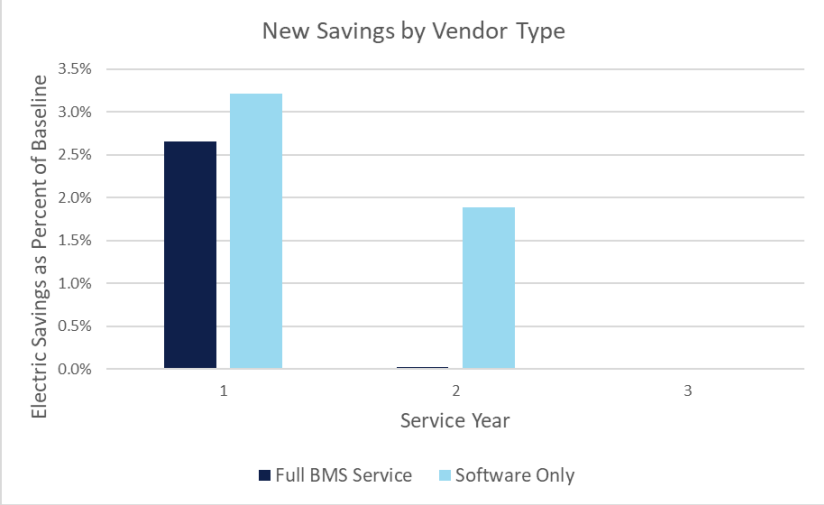
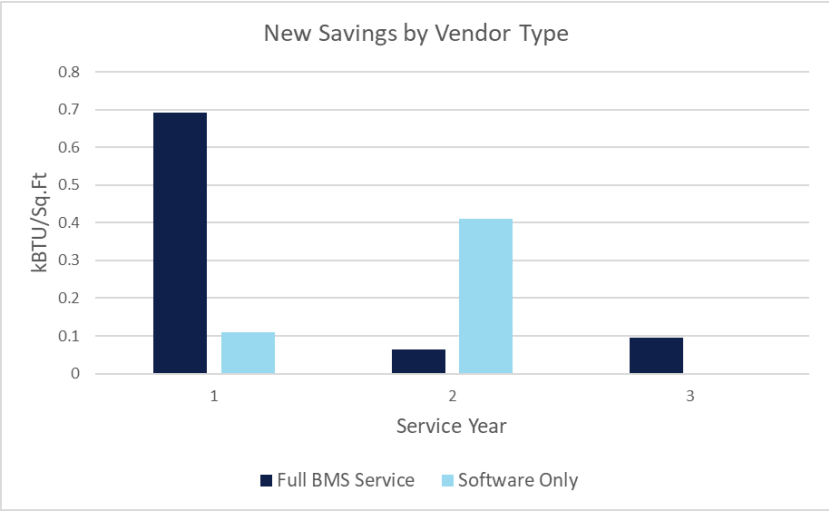


Figure B- 1 above shows electric energy savings as a percent of baseline from newly installed measures for each service year. For both systems, Year 1 achieved more savings than subsequent years which is to be expected given that the systems are likely capturing easy-to-implement measures at first.

Figure B- 2 Incremental heating fuel savings over time by vendor type



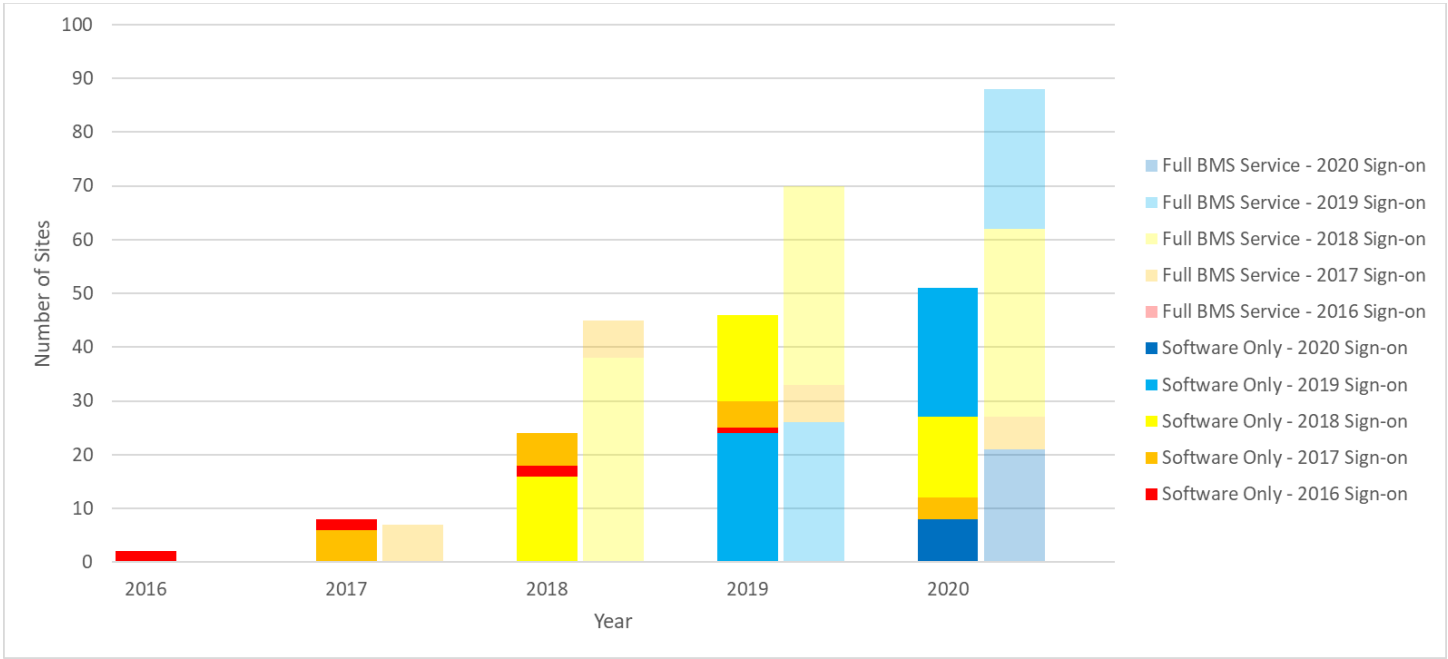
It is also important to mention that the values presented in the figure represent the population and were derived by expanding the sample. Table B- 1 shows the composition of both the sample population by count and percent baseline energy use.

Table B- 1 Distribution of vendor types in both sample and population

Metric	Software Only		Full BMS Service	
	Sample	Population	Sample	Population
Project Count (% of Total)	62	147	118	321
Baseline Energy Use (% of Total)	85%	86%	15%	14%
Square feet (% of Total)	70%	75%	30%	25%

Furthermore, the evaluation team examined the sample to determine customer engagement by vendor type. The evaluation team understands the pre-requisite consulting contract covers three years and can be extended for two additional years. Figure B- 2 below shows the persistence of customers by vendor type. Each color represents participants for a particular program year, whether or not that group is still engaged can be shown by the presence or absence of that color in subsequent years.

Figure B- 3 Customer engagement by vendor type



Savings by facility type

The evaluation team examined savings by all facility types in the population. The conclusion was that most savings are concentrated in two major facility types which are commercial office and multifamily. As a result, to draw a clear comparison and to appropriately represent the population, the evaluation team grouped all other facility types into one category.

The section that follows compares savings over time by commercial office, multifamily and other.

Figure B- 4 Incremental electric savings over time by facility type

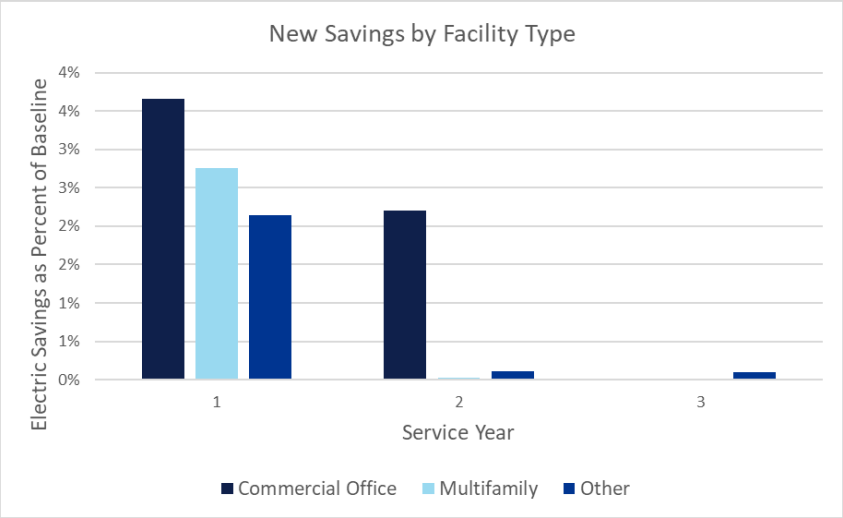
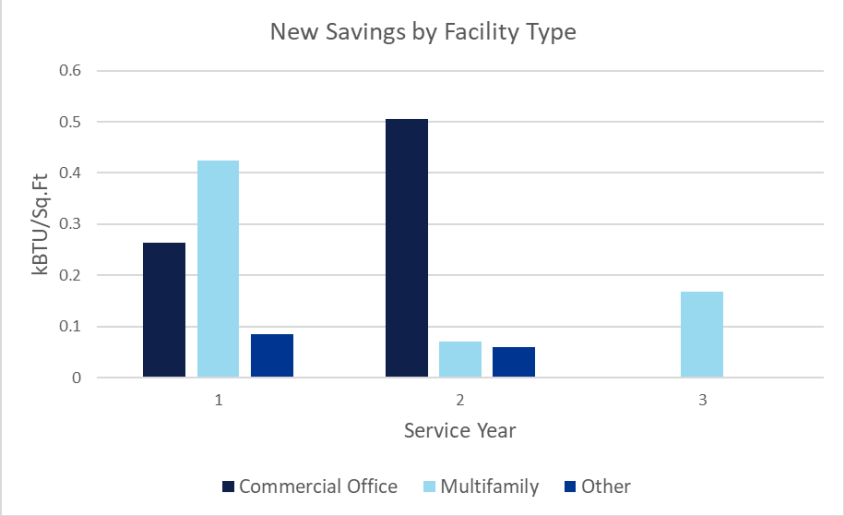


Figure B- 4 above shows that commercial offices managed to reduce the most energy with respect to baseline. It's important to note that 73% of the population baseline energy use is attributed to commercial offices which indicates could be a source of bias to the data presented above.

Figure B- 5 Incremental heating fuel savings over time by facility type



Savings by measure type

Finally, the evaluation team examined the sample to determine the distribution of savings by measure type. This could only be conducted on the sample level since measure information is not available on the population level.

Figure B- 6 and Figure B- 7 draw a comparison of installed and not installed electric energy savings by measure type.

Figure B- 6 Installed electric savings by measure-type

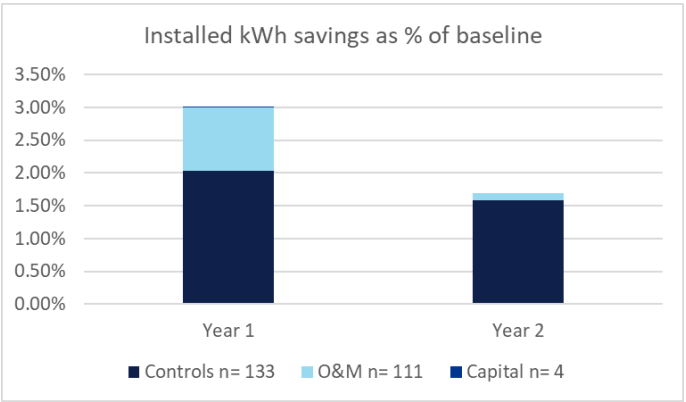
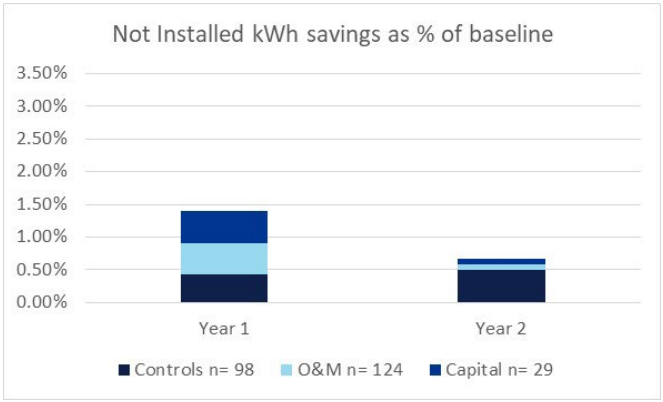


Figure B- 7 Not installed electric savings by measure-type



Similarly, Figure B- 8 and Figure B- 9 compare installed and not installed heating fuel energy savings by measure type.

Figure B- 8 Installed fuel savings by measure-type

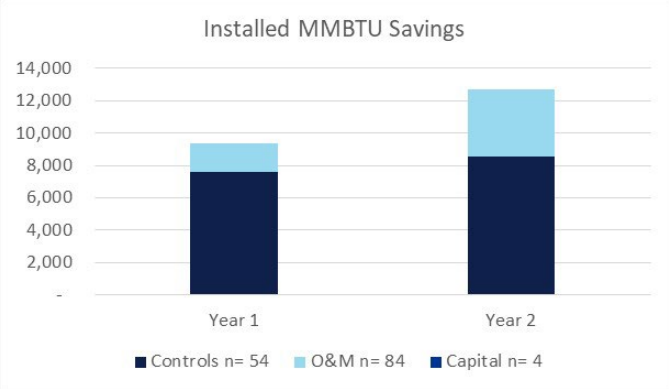


Figure B- 9 Not installed fuel savings by measure-type