Energy Trust of Oregon 2021 Existing Buildings Evaluation Final Report

Submitted to

Cody Kleinsmith
Energy Trust of Oregon
421 SW Oak St., Suite 300
Portland, OR 97204

Submitted by

Jeremy Stapp
Faith DeBolt
Dan Bertini
SBW Consulting, Inc.
2820 Northup Way, Suite 230
Bellevue, WA 98004

September 28, 2023

sbw

ENERGY + WATER + EFFICIENCY

Table of Contents

List of Figures

List of Tables

Executive Summary

This report details the findings and methodology we used to evaluate savings claimed in 2021 by Energy Trust of Oregon's Existing Buildings (EB) programs. We cover the following program tracks in this report: Lighting, Standard, Custom, and Strategic Energy Management (SEM). Findings and results are summarized below and covered in more detail in subsequent sections.

Evaluation Objectives

The objectives of this evaluation were to:

- Develop reliable estimates of EB program gas and electric savings for 2021 to establish realization rates. Realization rates are provided separately for SEM and non-SEM measures. This information will be used for future program savings projections and budget developments.
- Develop estimates of electricity and gas demand savings at the program track level (excluding SEM) and for the program overall (excluding SEM).
- Report observations from the evaluation and make recommendations to help Energy Trust understand substantial deviations from claimed savings, and to improve ex ante savings estimates and the effectiveness of future engineering studies and impact evaluations of Existing Buildings projects.

Methodology Overview

We used the process shown in Figure 1 to conduct the evaluation. A brief overview of each aspect of the process follows.

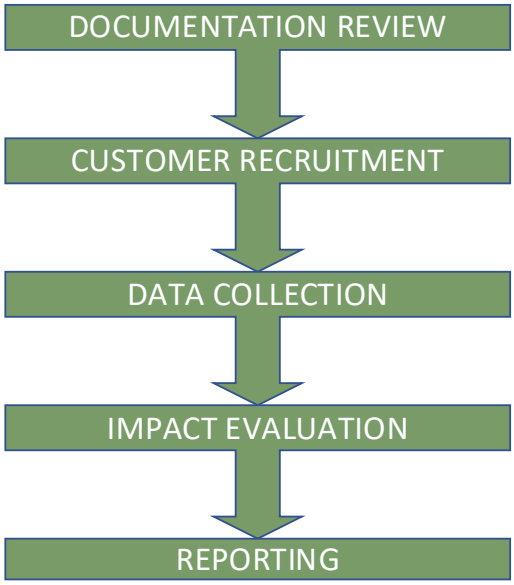

Figure 1: Evaluation Process Steps

Documentation Review

For each sampled site, we reviewed project documents supplied to us by Energy Trust. We determined whether or not sufficient documentation was present to perform an evaluation for the site. We requested any missing information from Energy Trust. We then determined the site treatment that would be used to perform data collection. Site treatments considered included desk review, virtual site visit, or physical site visit (although we did not conduct physical site visits for 2021—this is discussed further below).

Customer Recruitment

The Program Management Contractor (PMC)[1] first contacted sampled customers for Custom, Lighting, and Standard sites and the Energy Trust SEM coaches contacted SEM sites. Once a site was contacted by the PMC, we contacted the customer via email or phone, asking them to participate in the evaluation and providing specific details about data needed/interview questions.

If the customer was not willing or able to participate, we considered selecting a replacement site if the project timeline supported it. If the customer was unresponsive after two attempts at communication by us, we asked the PMC to attempt further contact. If contact was still unsuccessful at this point, we considered a final attempt by the PMC and then either selected a replacement site, dropped the site from the sample, or performed a no-contact review.

Note that we experienced a number of recruitment challenges for the 2021 evaluation (similar to the 2020 evaluation), due mainly to ongoing impacts related to COVID-19. These challenges resulted in fewer projects evaluated as compared to the original sample. Relative precisions were also affected, but final results were still found to be statistically relevant at the program level. Note that precision for some tracks however was low. Implications of the low precision in results for these tracks is discussed further throughout the report.

Data Collection

For Standard and Lighting projects, we used measure-specific data collection plans developed by us and approved by Energy Trust. In most cases, these data collection plans were specific to the Measure Approval Document (MAD) governing the measure.

For Custom and SEM sites, we developed site-specific data collection plans. The overall structure of these site-specific plans was approved by Energy Trust. Additionally, Energy Trust's SEM team reviewed and approved a sample of site-specific plans at random.

We used data collection plans to conduct data collection for each site.

For desk reviews, data collection consisted of an email exchange and/or phone call with the customer to confirm data entries.

[1] PMC for the Custom and Standard Tracks in the 2021 program year was TRC, Inc. The PMC for the Lighting Track in 2021 was CLEARResult, but TRC also provided the first contact for Lighting Track sites.

For virtual site visits, we set up a specific time with the customer to go over the project installation and any data needed. This included the customer taking photos, sending trend data, or if necessary, viewing the installed equipment with the evaluation engineer on the phone.

In general, we found that a virtual site visit was the most efficient, allowing us the time and access necessary to obtain the required data, but without the additional time and resources required of a physical site visit. Given the mix of measures that was sampled for the 2021 evaluation, we decided not to conduct physical site visits as we were able to obtain all data needed to support an acceptable level of rigor via either desk review or virtual site visits.

Impact Evaluation

We developed evaluation workbooks to be completed by the evaluation engineer for each project. In these workbooks, we recorded all findings, estimated energy savings at the site, calculated the project level realization rates, and documented reasons for any differences between claimed and evaluated savings.

The evaluation engineer used project documentation and assembled data from the site to complete the evaluation workbook. A more senior engineer then performed a quality control (QC) check of the site evaluation. Once the QC check was completed, the results for the site were considered final.

Reporting

We developed a system to report results at the sample and population level. This included a project database containing entries for each completed project evaluation and a series of scripts to combine results, extrapolate them to the population level, and to report out various aspects of the evaluation.

Evaluation Results

This section presents a brief summary of the results. We provide a summary of results of the 2021 evaluation followed by a discussion of historical results as compared to 2021. A more detailed treatment of findings and results is presented in subsequent sections.

Summary of 2021 Evaluation Results

Table 1 and Table 2 show a summary of evaluation results by program track for each fuel (electric and gas). Program tracks are grouped by Capital (non-SEM) and SEM. This includes number of measures/projects for which evaluations were completed and total energy savings for both the population and sample, resulting realization rates, and relative precision of each track in the sample.

For the SEM track, we present evaluation results from two methodologies: The Savings Rate Table (SRT) approach used by the program, and the PTT (Performance Tracking Tool) modeling approach. Note that while both approaches were evaluated, we used the PTT results

for the final evaluated savings for the reasons discussed further in this report. As PTT was used as our final result, it is the primary focus of the results discussions that follow. We do however discuss the SRT approach as well and how it might be improved upon so that results align better with realized (modeled) savings.

Overall, we calculated the electric realization rate to be 0.93, with the Standard track being the highest (1.06) and the Custom track the lowest at 0.66. Note that the relative precision for both SEM and Custom tracks was low. The low precision for Custom was the result of a low customer response rate for this track and hence a lower number of evaluated projects. For this reason, we consider this result uncertain.

The low precision for SEM was the result of a high degree of scatter between claimed (SRT-based) savings and evaluated (PTT-based) savings. This is an important observation as the low precision is not reflective of the PTT method itself but is more a reflection of the accuracy in the savings estimates between the PTT and SRT approaches. Although the SRT method yielded a higher precision in the realization rate, comparison with the PTT approach at the project level indicates that the PTT method is the more accurate approach to calculating project-level savings.

Table 1: 2021 Electric Energy Savings and Realization Rates[2]

Program Track	Electric Frame							
	Unique Measure Count	Unique Project Count	Evaluated Project Count	Population Claimed Savings (kWh)	% Savings Sampled	Population Evaluated Savings (kWh)	Realization Rate	Relative Precision @ 90% Confidence
Standard	405	318	31	16,858,042	41%	17,848,238	1.06	13.3
Custom	247	177	10	26,913,894	10%	17,737,402	0.66	34.5
Lighting	3,886	1,149	25	54,948,182	13%	56,060,433	1.02	10.0
Capital	**4,538**	**1,644**	**66**	**98,720,118**	**17%**	**91,646,073**	**0.93**	**9.4**
SEM - SRT	**449**	**449**	**27**	**12,701,358**	**32%**	**10,634,142**	**0.84**	**15.7**
SEM - PTT	**449**	**449**	**27**	**12,701,358**	**32%**	**11,492,829**	**0.90**	**67.9**
Grand Total w/ SEM SRT	**4,987**	**2,093**	**93**	**111,421,476**	**19%**	**102,280,215**	**0.92**	**8.6**
Grand Total w/ SEM PTT	**4,987**	**2,093**	**93**	**111,421,476**	**19%**	**103,138,902**	**0.93**	**11.3**

We calculated the overall gas realization rate to be 0.89 (Table 2). The Custom track had the highest realization rate at 0.99, with the SEM track being the lowest (0.56). Note again the lower relative precision for both SEM and Custom. As with the electric SEM and Custom

[2] The following rows in this table were used to generate SRAFs: Standard, Custom, Lighting, SEM-SRT.

results, the causes for low precision were lower number of completed project evaluations for Custom, and large scatter in site-level realization rates for SEM.

Table 2: 2021 Gas Energy Savings and Realization Rates[3]

Program Track	Gas Frame							
	Unique Measure Count	Unique Project Count	Evaluated Project Count	Population Claimed Savings (Therms)	% Savings Sampled	Population Evaluated Savings (Therms)	Realization Rate	Relative Precision @ 90% Confidence
Standard	172	140	26	968,411	39%	925,691	0.96	11.8
Custom	149	114	15	1,021,437	24%	1,007,478	0.99	25.5
Capital	**321**	**254**	**41**	**1,989,848**	**31%**	**1,933,169**	**0.97**	**14.4**
SEM - SRT	**379**	**379**	**26**	**481,830**	**30%**	**471,415**	**0.98**	**10.2**
SEM - PTT	**379**	**379**	**26**	**481,830**	**30%**	**267,565**	**0.56**	**85.1**
Grand Total w/ SEM SRT	**700**	**633**	**67**	**2,471,678**	**31%**	**2,404,585**	**0.97**	**11.8**
Grand Total w/ SEM PTT	**700**	**633**	**67**	**2,471,678**	**31%**	**2,200,734**	**0.89**	**16.4**

Figure 2 and Figure 3 show in graphical form the population-level realization rates by domain. Plotted circles within each domain represent sampled measures. The size of each circle represents the magnitude of claimed savings. Realization rate is shown by the circle's position on the y-axis. The error band is shown in red around the domain's realization rate (depicted as a red dot). The error band is a graphical representation of the relative precision of the domain's realization rate.

In Figure 2, to improve scalability we removed two small saver projects with high realization rates from the graph. These two projects were in the Standard domain and had the highest realization rates on the electric side. They are discussed further in Findings below. We identified a number of projects on the electric side with realization rates of zero, most of which were found in the SEM domain (PTT-based results) with a few in the Custom domain. Realization rates were more varied for the Custom track and were at or near 1 for the Lighting and Standard tracks. Note that error bands are larger for domains with more variance in project realization rates. Note also the large scatter in realization rates for the SEM track, which was the primary reason for low precision in this track.

[3] The following rows in this table were used to generate SRAFs: Standard, Custom, SEM-SRT.

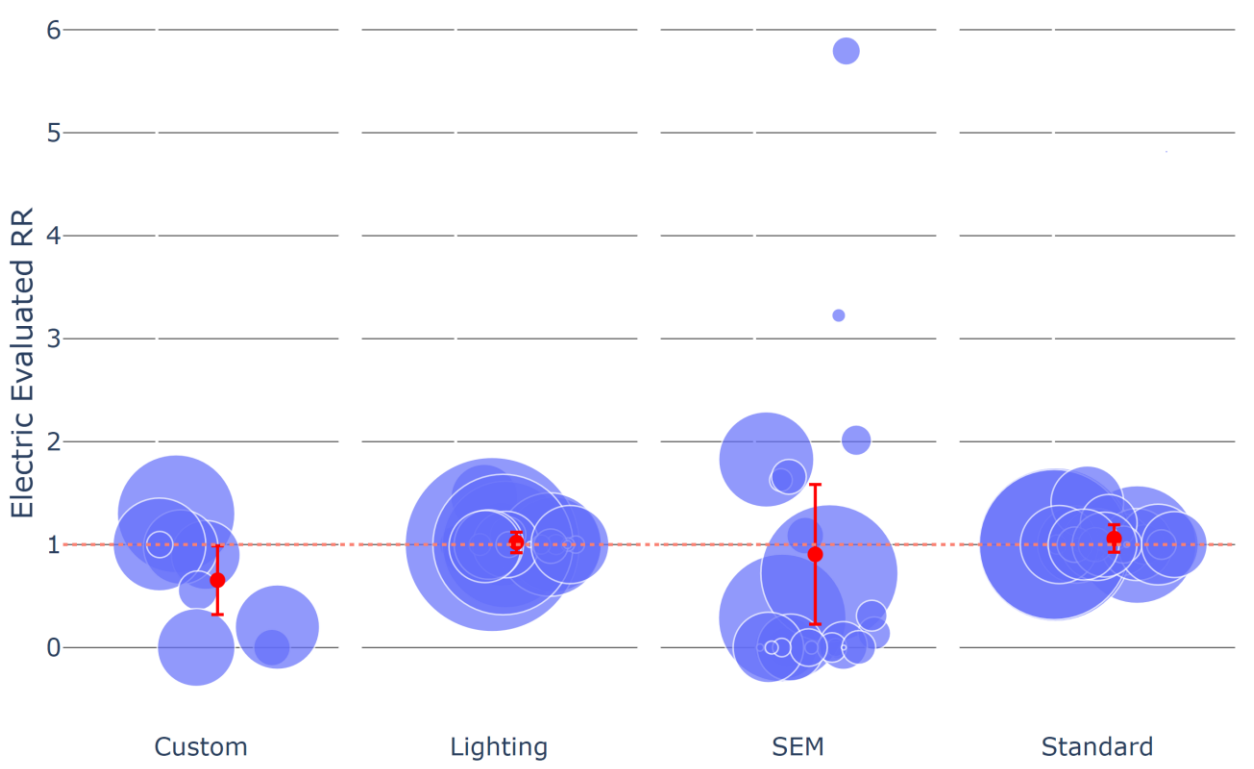

Figure 2: Electric Energy Savings Realization Rates by Domain

Figure 3 shows results for the gas sample. We removed one outlier with a realization rate above 6 from the SEM domain to improve scalability of the graph. Realization rates of zero were found mainly in the SEM track (PTT-based), with a few in the Standard domain and one in the Custom domain. Similar to the electric sample, error bands for domains with larger variability in project realization rate were larger. Note also (as with electric) the large scatter in SEM gas results, which was the primary driver of low precision for this track.

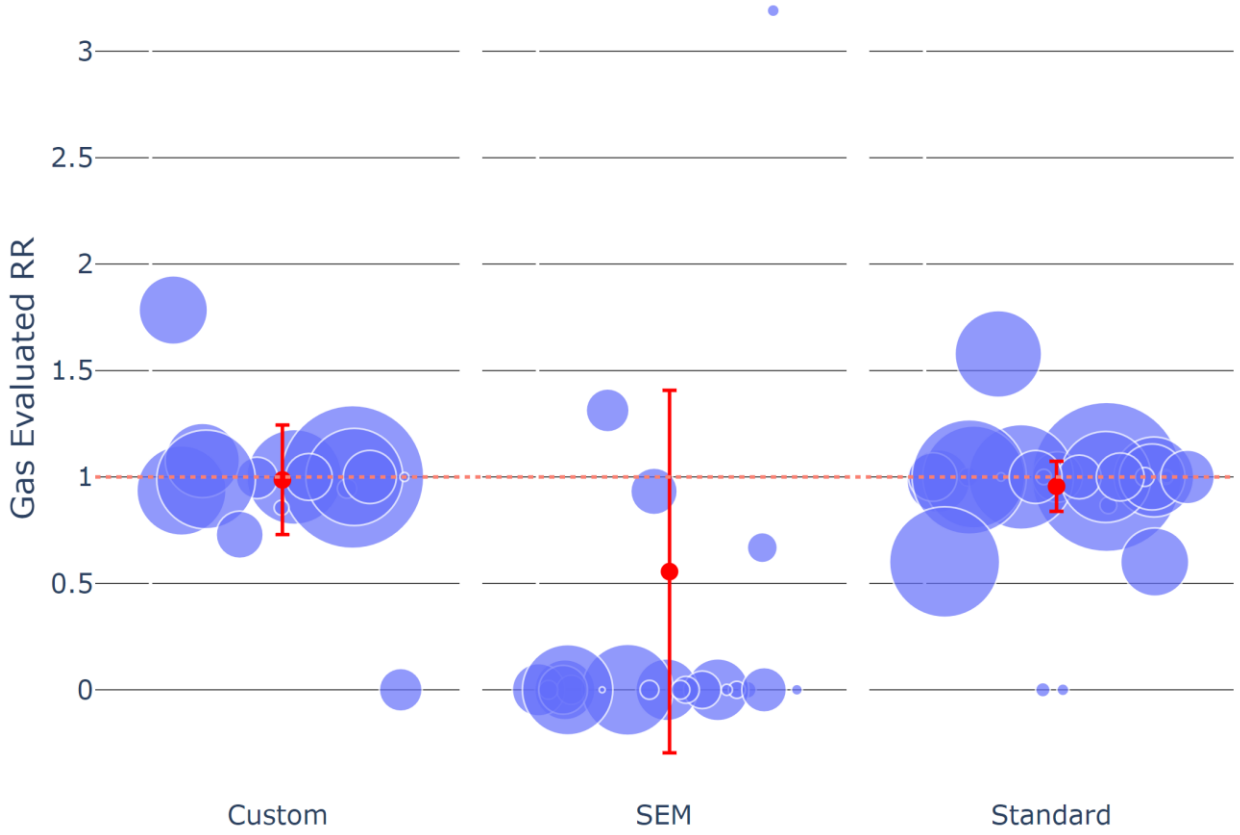

Figure 3: Gas Energy Savings Realization Rates by Domain

Table 3 and Table 4 show a summary of kW and gas demand savings results by program track. Demand savings were calculated for all non-SEM evaluated projects. To calculate electric and gas demand savings, we used Energy Trust-supplied peak demand factors per load profile for non-hourly electric and gas calculations (Standard and Lighting) and peak period definitions for hourly electric (Custom) calculations. We evaluated electric demand savings for both the summer and winter peaks. We evaluated gas demand savings as a single annual value.

Demand savings realization rates are shown together with energy consumption savings realization rates for comparison. Demand savings realization rates varied mainly due to variance in energy consumption savings. Custom had the smallest realization rates for both gas and electric, due primarily to changes in energy savings during peak periods.

Table 3: Comparison of Electric Energy and Demand Savings Realization Rates

Program Track	kWh RR	Summer kW RR	Winter kW RR
Standard	1.06	1.10	1.00
Custom	0.66	0.77	0.84
Lighting	1.02	1.00	1.00
Overall	**0.93**	**0.98**	**0.99**

Table 4: Comparison of Gas Energy and Demand Savings Realization Rates

Program Track	Gas Consumption RR	Gas Demand RR
Standard	0.96	0.97
Custom	0.99	0.39
Lighting	NA	NA
Overall	**0.89**	**0.90**

Discussion of Historical Results

The following figures display historical evaluation results (total energy savings and realization rate) by year from 2008 to 2020 together with 2021 (the results of this evaluation). Results are shown separately for electric and gas, and for SEM (PTT-based results are shown for 2021) and non-SEM program tracks.

The figures indicate the following:

- Non-SEM electric savings (Figure 4) shows an overall upward trend in total energy savings over time but with a significant drop in 2020 and slight increase in 2021. The realization rate has been relatively steady in recent years at close to 1, but did drop somewhat in 2021.

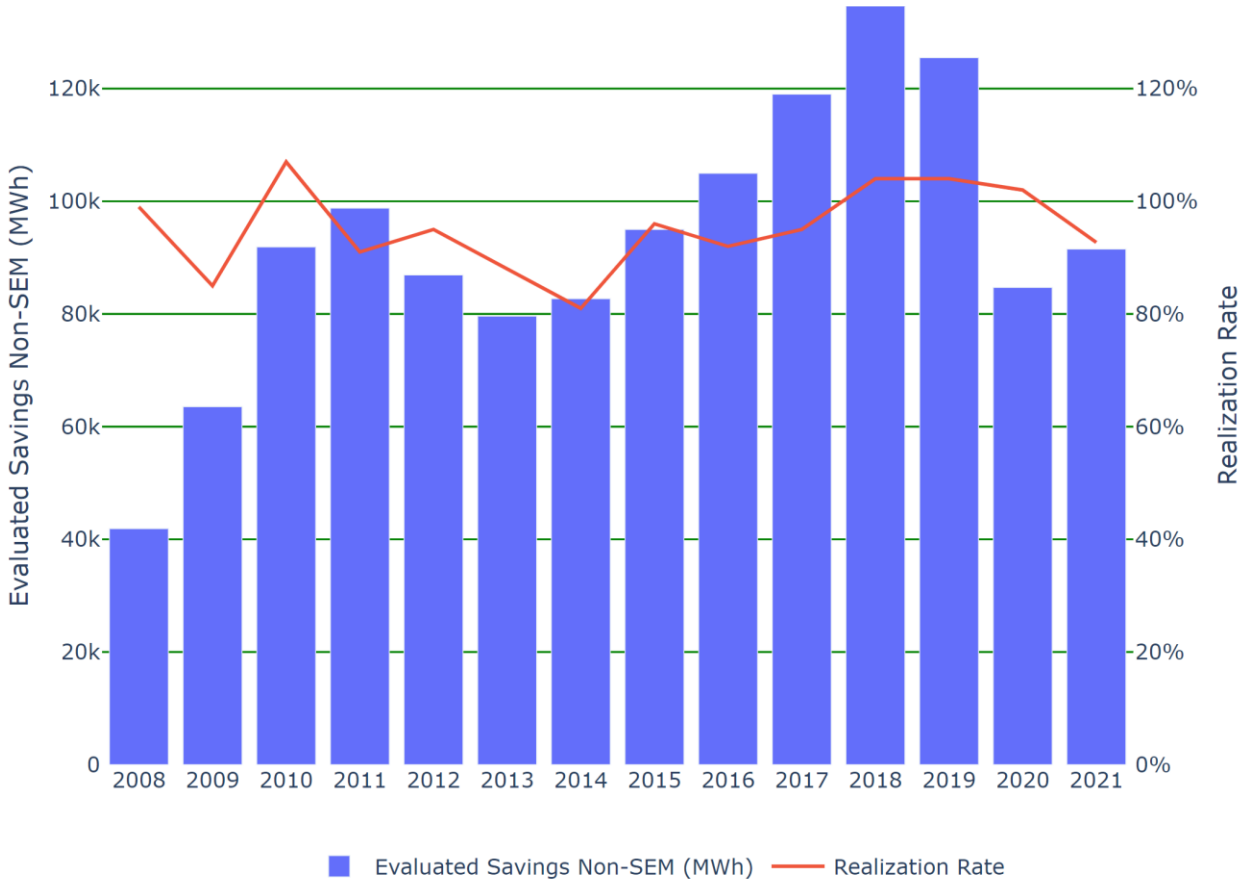

Figure 4: Historical Non-SEM Electric Energy Savings Evaluation Results

- The non-SEM gas savings trend was somewhat flat in recent years but increased significantly in 2021 (Figure 5). Historically, the realization rate was lower than non-SEM electric but trended upwards between 2018 and 2021.

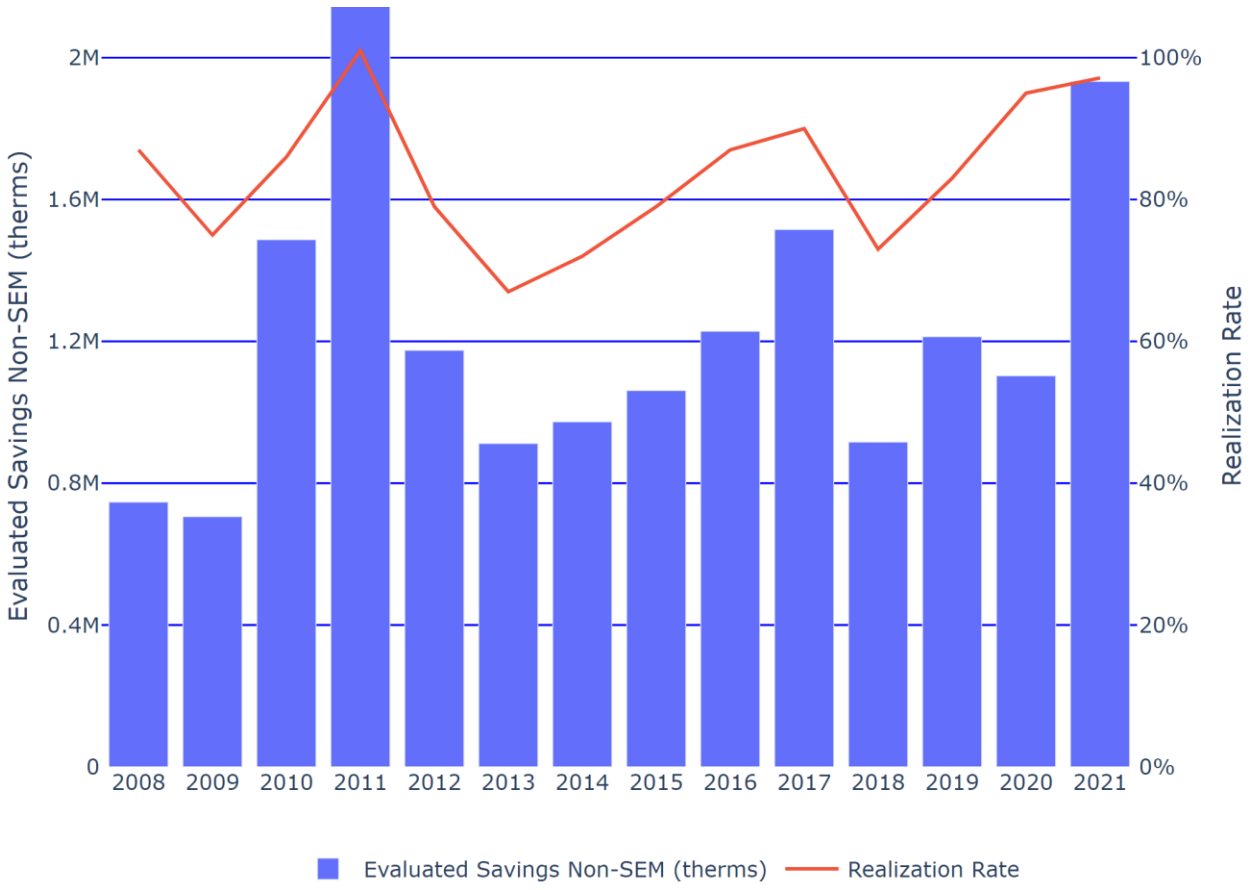

Figure 5: Historical Non-SEM Gas Energy Savings Evaluation Results

- SEM electric savings (Figure 6) showed a significant increase after 2017 with a steady realization rate near 1 between 2014 and 2019. 2020 saw a notable decrease in realization rate, with a rebound in 2021. A downward trend in savings is also apparent between 2018 and 2020, but again with a rebound in 2021.

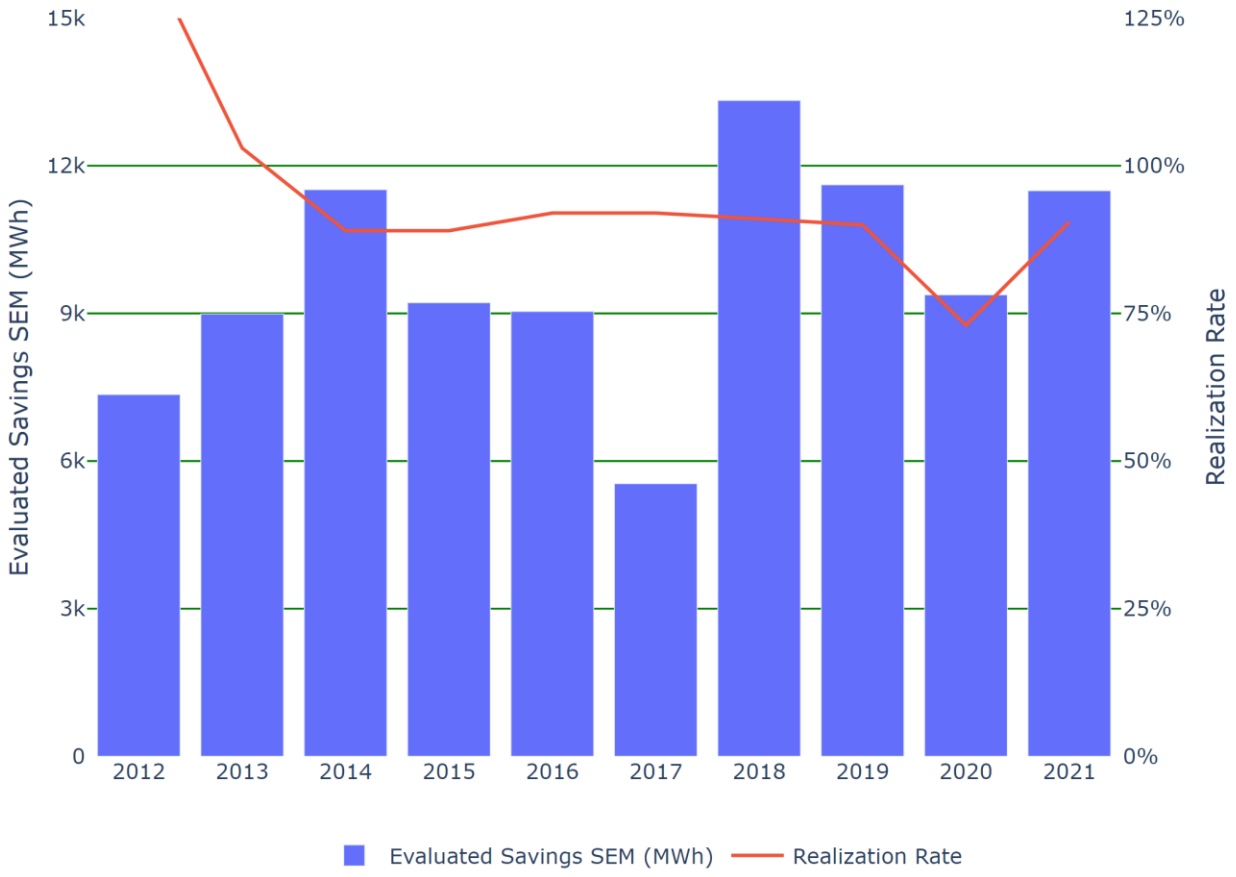

Figure 6: Historical SEM Electric Energy Savings Evaluation Results

- SEM gas savings (Figure 7) does not show a clear trend over time. The realization rate has been somewhat volatile with an average near 1. The realization rate remained near 1 for 2020, but dropped in 2021 along with savings claimed.

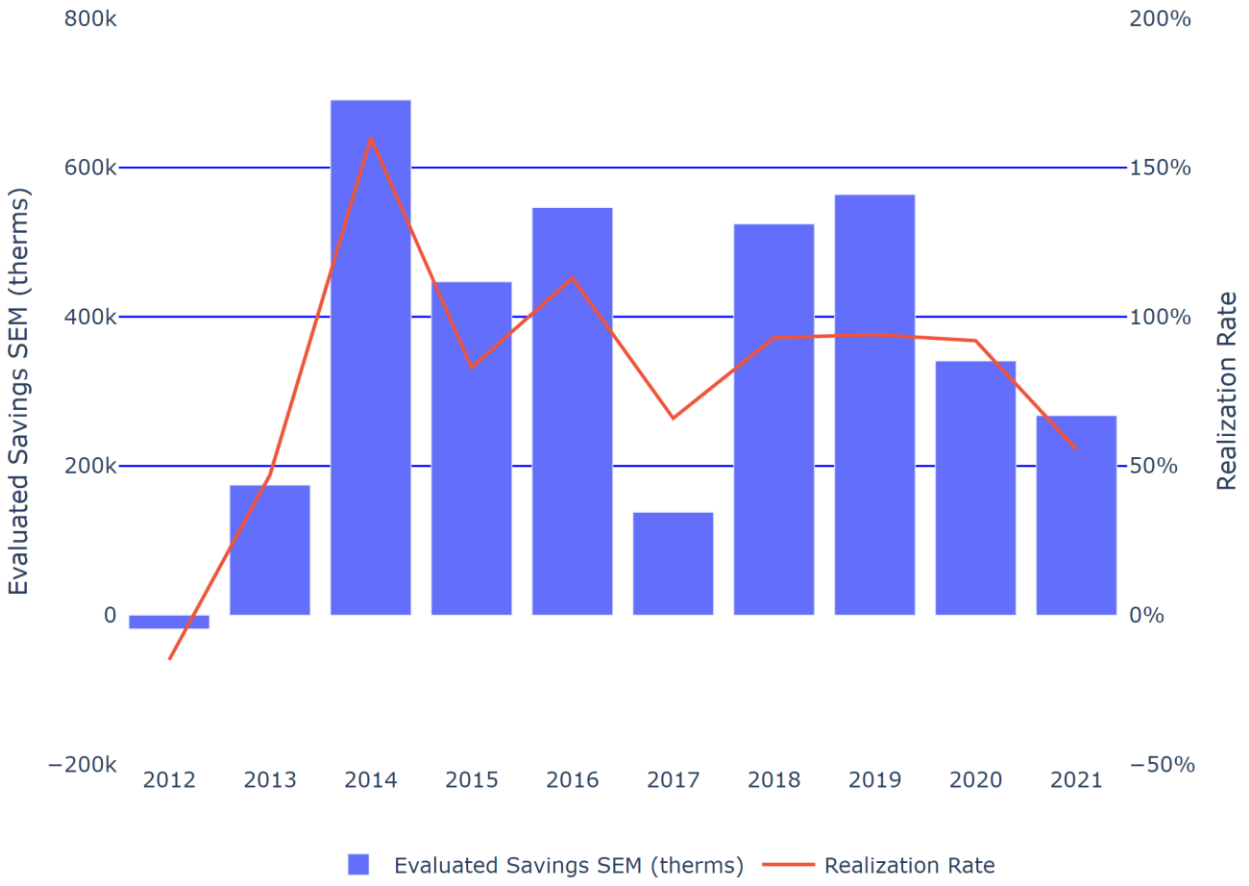

Figure 7: Historical SEM Gas Energy Savings Evaluation Results

Evaluation Observations/Recommendations

The purpose of this task was to document observations made about the program during the course of the evaluation and to make recommendations to help Energy Trust improve the effectiveness of the program and the accuracy of the expected savings in future program years.

The data collection and analysis procedures described in detail in the Evaluation Methodology section support documenting observations in a standardized way to facilitate qualitative and, in some cases, quantitative findings on how well the program is operating.

We made the following observations in 2021 during the evaluation process. For each observation, we provide a recommendation for improvement.

Custom Track

- **Custom Energy Models:** During the course of our model reviews, we noted that model calibration was inadequate for 12 of 16 models. This issue is further detailed in section 3.2.5. Note that most of these models were developed under the previous PMC.

  Recommendation: We recommend that the PMC check to ensure that model calibration is being emphasized in technical reviews and that any explanations given by the ATAC for poor calibration are thoroughly reviewed to ensure that PMC calibration guidelines are adhered to. If a model is unable to be calibrated following these guidelines, we recommend that the model not be accepted and that a different calculation approach be used for the project.

Standard Track

- **Standard Track MAD 47 (Insulation) and MAD 68 (Cooler Doors) Existing Conditions Documentation:** Existing conditions for insulation (MAD 68) and cooler doors (MAD 47) were not documented in some cases.

  Recommendation: Since the existing condition is a key input for measure savings/incentives, we recommend carefully checking this documentation before approving applications, and ensuring that the customer understands that this documentation is required per the MAD.

Lighting Track

We did not have specific recommendations for the Lighting track. We did not identify any issues with this track that we felt would warrant a recommendation.

SEM Track

- **SEM PTT Method:** For this evaluation, we used the PTT method to calculate realized savings. The overall sampling precision of the SEM realization rate based on the PTT results is very low due to the scatter in site-level realization rates and is not reflective of the accuracy of this method for estimating savings at the site level. At the site level, we were able to successfully model savings at each site and achieve acceptable model fit metrics with adjustments for NRAs such as those associated with the pandemic.
- **Recommendation:** We believe that this method, which utilizes actual site level data, is the preferred approach and can be successfully implemented during periods such as the recent pandemic.
- **SEM SRT Method:** Our analysis indicates that the program did a good job implementing the SRT method for claiming savings but that the SRT method itself could be improved if it is used in the future.

Recommendation: If the SRT method is to be used in the future, we recommend basing the site-level savings not wholly on the historical savings rates for the program, but on a combination of the historical savings rates for the program and, if available, the historical savings rates for the site, or the participant. For example, we found nine PTT models in their second reporting year or greater where savings had never been realized in any of the previous reporting years. We found that seven of those nine did not have savings during the evaluation year, either, yet by the SRT method all nine of those sites had savings.

Additionally, if the program uses a similar method in the future, the evaluator should be informed that a higher sample size may be necessary to achieve desired precision targets since the claimed savings may vary significantly from the realized savings.

Finally, we recommend that the engagement factor (EF) be studied more closely to quantitatively assess the impact that the number and types of "qualifying projects" has on savings. Moreover, we would recommend assessing the engagement factor at the building level, if possible, rather than at the customer level.

- **SEM Monthly vs Interval Billing Data:** The SEM program has historically used monthly billing data when using the PTT modeling approach.

  Recommendation: We strongly recommend using interval data (i.e. daily data) if at all possible for regression-based savings models. Interval data usually results in much smaller uncertainty as compared to monthly data. Our understanding is that interval data is difficult to obtain from utilities at the time of this report writing. Utilities would benefit greatly from sharing this data as it would increase model accuracy and therefore aid in increased energy efficiency.

General Recommendations

- **Low Customer Recruitment/Participation:** As detailed in 3.1.1, customer recruitment levels and cooperation were low in 2020, and despite greater focus on the recruitment process, were also low in 2021.

  Recommendation: We believe that evaluation participation could be improved by increasing communication with the customer regarding the evaluation process, and continuing communication with them throughout the measure implementation process so that it is repeatedly clear what the expectations are if they are contacted for an evaluation. This would also include re-engaging the customer on this process if the customer contact person changes to avoid situations in which the site contact is unaware of the evaluation process. It may also be good to emphasize to the customer that participation in the evaluation process is technically required as part of their contract with Energy Trust. Finally, since this seems to be a persistent issue, it might benefit from further study perhaps as part of a process evaluation. This would help ensure that any barriers to participation are thoroughly identified.

- **Low Precision in Track Level Results:** Precision was much lower than targeted in both the Custom and SEM tracks.

  Recommendation: Precision was affected by low customer participation (lower final completed sample) in the Custom track and high variability of results (claimed savings vs evaluated savings) in the SEM track. One way to mitigate these factors in the future would be to use the relative precision and site completion rates (completed sites/originally sampled sites) from past evaluations to drive future evaluation sample sizes. A higher sample size for example would result in more sites being completed (in the case of low customer participation) and better statistical results (in the case of high scatter in results). This approach would require a higher evaluation budget to handle the increased effort for recruiting and evaluating more projects.

Memo

To: Board of Directors

From: Cody Kleinsmith, Evaluation Project Manager
Oliver Kesting, Sector Lead – Commercial
Patrick Urain, Sr. Program Manager – Commercial
Kathleen Belkhayat, Program Manager – Commercial

cc:

Date: November 3rd, 2023

Re: Staff Response to the 2021 Existing Buildings Impact Evaluation

The 2021 Existing Buildings Impact Evaluation assessed the performance of projects claimed in the 2021 program year in the program's four main tracks: Custom, Lighting, Standard and Strategic Energy Management (SEM). The program management contractor (PMC) for 2021 was TRC, with delivery of the lighting track subcontracted to CLEAResult (program delivery contractor, or PDC). For both TRC and CLEAResult, this was their first year under their respective PMC and PDC contracts, and many projects that were closed out were initiated by the previous contractors. The results of the evaluation show the program performed well in 2021 despite ongoing challenges due to the COVID-19 pandemic. While 2021 savings realization rates were lower than some previous years, particularly for gas, overall program realization rates were high, at 93% for electricity and 89% for natural gas.

Due to changes in building occupancy and operations during the pandemic, SEM continued to use an alternative method of claiming savings. This method claimed savings using the program's historical savings rate, customer participation level, year in the program, and other factors rather than using estimated savings from energy models. This evaluation determined two realization rates for SEM. The first, Savings Rate Tables (SRT) realization rate, determined how well the program implemented this alternative method and resulted in realization rates of 84% for electric and 97% for gas. The second, Performance Tracking Tool (PTT) realization rate, replicated the standard (non-COVID year) modeling approach to evaluate savings. This method returned realization rates of 90% for electric and 56% for gas. The overall realization rate of the PTT method had low precision, primarily due to the large variation of realization rates in individual projects across the evaluated sample. This low precision makes it difficult to draw conclusions across the SEM track from this sample of evaluated projects.

The program updated the alternative SEM savings methodology prior to the 2021 program year, and these large differences in SRT and PTT realization rates indicate there are more changes needed to fine tune the SRT approach if the program plans to use it again in the future. In 2023, the program transitioned back to using energy models to claim savings. If a large macroeconomic disruption similar to the COVID-19 pandemic occurs in the future, the program will consider if the SRT approach is suitable enough to claim reliable savings compared to modeling. If the SRT method is put in place again, the program will incorporate changes recommended in this evaluation to strengthen its ability to reliably claim savings.

The evaluation found custom track savings models developed by the previous PMC and closed out and claimed by TRC in 2021 had more errors in model calibration than projects with models developed, closed out and claimed entirely by TRC. The Existing Buildings program will explore ways to ensure models

developed near or during PMC contract conclusion dates are developed with the same rigor as models developed throughout other program contract years.

At the beginning of 2021, implementation of Business Lighting (which also serves industrial customers) was transferred to CLEAResult. The lighting track returned a realization rate of 102% with no areas that could be identified as systematic places for improvement. The standard track, delivered by TRC as part of their PMC contract, also saw high realization rates at 106% for electric and 96% for gas with no areas where systematic improvements could be made.

As with the 2020 Existing Buildings Impact Evaluation, evaluators encountered challenges in recruiting participants to provide information, despite adjustments to recruiting methods. Issues of turnover or loss of participant staff resulting from the effects of the COVID-19 pandemic have remained, and more customers stopped participating in the evaluation mid-way through the data collection process. Evaluation staff will consider changes to evaluation scopes and budgets to account for lower evaluation recruitment rates, while continuing to explore more effective strategies for recruiting participants into evaluation.

Due to stable findings for the program realization rates over several years – outside of SEM, which has returned to using a modeling approach in 2023 – Energy Trust will not conduct an impact evaluation for the Existing Buildings' 2022 program year. The 2023 Existing Buildings Impact Evaluation will begin in 2024.

1 Background

1.1 Existing Buildings Program Overview

In 2021, the Existing Buildings (EB) program had four main program tracks:

- **Custom.** Custom track project energy savings are estimated through energy studies conducted by Allied Technical Assistance Contractors (ATACs). These studies involved engineering calculations (e.g., Excel workbooks) and/or energy simulation modeling (e.g., eQUEST, TRACE700, etc.).
- **Lighting (prescriptive and semi-prescriptive).** Lighting track measures were installed directly by trade allies. These measures (with the exception of Custom Lighting/Controls measures) are specified in Energy Trust Measure Approval Documents (MADs).
- **Standard (prescriptive).** Standard track measures used savings estimates from reliable sources (including the Regional Technical Forum, ENERGY STAR, and others). Similar to Lighting track measures, Standard track measures are specified in Energy Trust MADs.
- **Strategic Energy Management (SEM).** SEM provides tools and education to help businesses understand their building energy use and identify and eliminate energy waste. SEM engagements last about a year and participants have the option of re-enrolling annually. SEM savings are typically estimated based on a top-down analysis of building-level energy use and do not include savings from capital measures completed at the site through other program tracks during the SEM engagement. Due to effects of COVID-19 on building operations, for 2021, the program calculated site-level incremental savings in a semi-prescriptive fashion by multiplying the baseline energy usage by a nominal historical saving rate (SR%) and engagement factor (EF).

The EB program maintained a few other tracks and pilots during the 2021 program year. These represented a small portion of program participants and savings, so the Energy Trust evaluation manager excluded them from the scope of this evaluation.

Note that this report refers to all non-SEM projects as "Capital" projects. This includes the following:

- All projects completed at non-SEM participant sites.
- Projects completed at SEM participant sites but incentivized and claimed through a non-SEM track. We evaluated many capital projects completed at sampled SEM participant sites to adjust SEM project evaluation savings appropriately[4]. Some of these SEM capital projects happened to be in the non-SEM sample.

[4] Where incremental SEM savings were negative for the site, we did not evaluate the SEM-capital project if doing so could not have resulted in positive incremental savings. Neither did we evaluate the SEM capital projects for which the claimed capital savings were significantly less than the uncertainty of the total modeled savings for the year.

1.2 Evaluation Objectives

The objectives of this evaluation were to:

- Develop reliable estimates of EB program gas and electric savings for 2021 to establish realization rates. We provide statistically significant realization rates separately for SEM and non-SEM tracks. Energy Trust may use the results for future program savings projections and budget developments.
- Develop estimates of electricity and gas demand savings at the program track level (excluding SEM) and for the program overall (excluding SEM).
- Report observations from the evaluation and make recommendations to help Energy Trust understand substantial deviations from claimed savings, and to improve ex ante savings estimates and the effectiveness of future engineering studies and impact evaluations of Existing Buildings projects.

2 Evaluation Methodology

This section covers in detail the methodology and procedures used to evaluate each track (Lighting, Standard, Custom, SEM). We first developed a Sampling Plan to generate the sample for the evaluation. We then evaluated each sampled project following these steps:

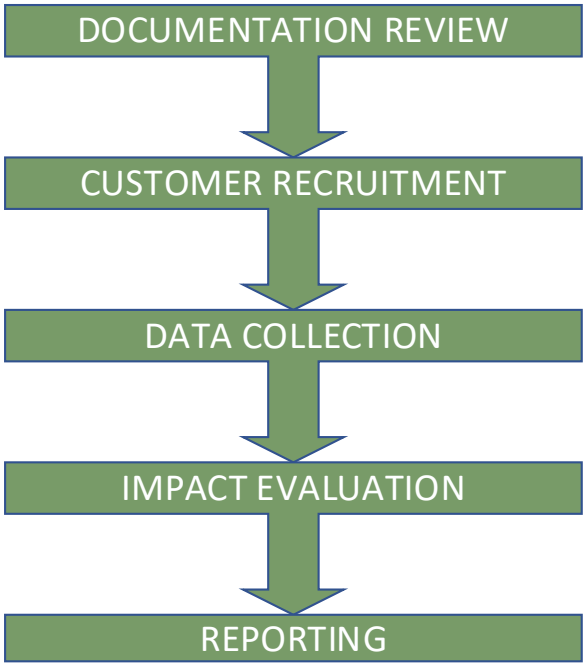

Figure 8: Evaluation Process Steps

2.1 Sampling Plan

We developed the sample design by first creating a sample frame from the 2021 program tracking data provided by Energy Trust. Energy Trust has multiple objectives for the sample design. One objective was to achieve reliable electric and gas savings estimates and realization rates separately for SEM and non-SEM measures. Another was to develop electricity and gas demand savings estimates at the program track level, excluding SEM.

Our plan for accomplishing the first objective was as follows:

- **Create separate lists of projects.** One list contained all projects with electric energy savings. The other contained all projects at sites with gas savings. The sample unit was by project, except SEM which was by site. A project may have multiple measures and/or both gas and electric savings. A site may have multiple projects, or a project may be across multiple sites, like SEM. After drawing the sample, we sorted the list by site to recruit by customer in an effort to minimize their burden from participating in the evaluation.
- **Certainty selections of large savers.** Within each fuel and domain, we identified the projects with the largest savings and selected those with certainty. This ensured that the sampled sites accounted for a large fraction of the total savings claim for the domain.

- **Sampling projects.** For the remaining projects in each fuel and domain, we developed an optimal stratification design. We defined strata based on the program estimate of savings using the Dalenius and Hodges method. We the applied the Neyman allocation method to determine an optimum sampling fraction for each stratum, referencing error ratios from recent Existing Buildings evaluations. For sampled SEM sites, we evaluated many capital projects implemented during the performance period[5]. We selected additional projects to allow for replacements, if needed.

We designed this sample plan to achieve 10% precision at 90% confidence in estimating savings and realization rates separately for SEM and non-SEM tracks by fuel.

Table 5: Sample Design

Track	Target Sample Size				Target Precision @ 90% Confidence	
	Electric	Gas	Total Unique Projects	Total Unique Measures	Electric	Gas
Lighting	30	-	30	168	9%	-
Standard Other	30	25	51	65	12%	11%
Custom	23	20	38	53	13%	13%
Non-SEM	**83**	**45**	**119**	**286**	**6%**	**9%**
SEM	**25**	**25**	**48**	**48**	**10%**	**9%**
Overall	**108**	**70**	**167**	**334**	**6%**	**7%**

The SEM sample comprised 27 unique SEM customers and 11 of those customers had two or more sites in the electric and/or gas sample. The intent was to allow us to provide evaluation feedback at the customer level, in addition to the site level.

For the second objective, we estimated electric and gas demand savings for the sampled projects, excluding SEM.

The final sample frame was used to initialize the program evaluation database and track each sampled project throughout the evaluation process.

2.2 Customer Recruitment

This section details procedures for recruiting customers to participate in the evaluation. The goal of the recruitment process was to recruit customers efficiently while avoiding undue hassle to the customer and minimizing dropped sites.

[5] Only SEM capital projects that were also sampled in other domains are included in evaluation results for their respective domains.

The recruitment process proceeded as follows:

- SBW provided Energy Trust and the Program Management Contractor (PMC) the final sample detailing all primary sites to be evaluated.
- The PMC reviewed the customer contact information for sampled sites in the program tracking database and made changes as necessary to ensure the best available contact information to support recruitment.
- SBW and Energy Trust drafted email introductions to be used by the PMC for the first contact with each site. This email established the importance of this work and Energy Trust/PMC/SBW role in performing the work. It also introduced SBW as the evaluation contractor.
- As projects were deemed ready for recruitment by SBW, the PMC used the email introduction provided by SBW to contact the sampled sites and cc SBW on the communication.
- SBW called or emailed the appropriate site contact to recruit the customer for the evaluation. The target for this communication was within one week of first contact by the PMC. At this stage we attempted to confirm that the customer was able to provide a person that was knowledgeable about the measure(s). If successful, for the sampled measures requiring site visits, we also attempted to determine the most appropriate contact for the site visit (virtual or in-person). This step was targeted to be completed within three weeks of first contact by the PMC.
- If the recruitment was not successful, we considered replacing the sampled case with another case from the same stratum in the sample. In some cases we performed a no-contact review (see 3.1.1). The recruitment was considered unsuccessful under the following conditions:
  - The customer refused to participate.
  - A person with adequate knowledge of installed measures could not be confirmed.
  - Reasonable access to the measure was needed but was not possible.
  - The customer did not respond to repeated communication attempts. In this case, SBW attempted communication at least twice. If the customer did not respond, we asked the PMC to attempt to contact the customer. If the customer did not respond to the PMC's attempt, we considered the recruitment unsuccessful. This represents a total of at least four contact attempts. We felt that this number of attempts balanced enough attempts with potential customer sensitivity to too many attempts.
  - In cases where recruitment was unsuccessful, we considered either exchanging the site with a replacement site or performing a no-contact review. If neither of these solutions was possible, we dropped the site without review or replacement. As the latter would result in a reduced sample, we discussed this with Energy Trust before proceeding.

We worked closely with Energy Trust and the PMC to avoid dropping and replacing sites to minimize non-response bias.

2.3 Documentation Review

We assigned a lead analyst and QC engineer to each sampled site. The lead analyst was given primary responsibility for both data collection and analysis of savings. The QC engineer was a more experienced member of the team responsible for advising the lead analyst and reviewing evaluation results for the project.

Energy Trust provided electronic copies of the project files. For custom measures, Energy Trust also provided electronic versions of the savings calculation spreadsheets, simulation input/outputs and other information to support the program savings estimate. The spreadsheets and simulation inputs/outputs were checked to ensure they were in a form that allowed us to easily reproduce the program estimate of expected energy savings. If projects had significant documentation issues or other issues that impeded the evaluation of savings, we worked closely with Energy Trust to decide on whether to continue to pursue evaluating the project, replacing with another project, or dropping the project from the evaluation without replacement – bearing in mind the potential to introduce bias by dropping/replacing a large number of sampled projects.

The documentation review had the following goals:

- Determine if there were any issues with the provided data such as missing documents, mis-matched model files, or other issues that would hinder the evaluation. We worked with Energy Trust staff, PMCs, Allied Technical Assistance Contractors (ATACs), and SEM coaches as necessary to resolve any issues found. If issues could not be resolved, we proceeded in the same way as for sites with unsuccessful recruitments (see recruitment process above).
- Identify data needed to complete evaluation for each track (Lighting, Standard, Custom, SEM). For Custom and SEM tracks, we developed site-specific data collection plans.
- For each site, we determined whether it would be evaluated via a desk review, virtual site visit, or physical site visit. See 2.3.4 for a complete definitions of site treatments.

Track-specific aspects of documentation review are discussed in the following sections.

2.3.1 Documentation Review: Standard/Lighting Tracks

The Standard and Lighting tracks include measures of lower complexity. Measures for these tracks are typically defined as Prescriptive (using deemed values to define savings) or Semi-Prescriptive (using a simplified calculator with a few inputs to calculate savings). With the exception of Custom Lighting/Controls measures, both Prescriptive and Semi-Prescriptive measures are defined in Energy Trust Measure Approval Documents (MADs).

The evaluation engineer was instructed to follow a documentation checklist developed based on the measure's related MAD. The evaluation engineer checked to ensure all information necessary to recreate savings inputs as required in the MAD were present (e.g., quantity/type of installed units, equipment efficiency, size, etc.).

2.3.2 Documentation Review: Custom Track

Custom track projects typically include whole building energy models (eQUEST, TRACE 700 or equivalent) or custom Excel spreadsheet calculations. They tend to be more complex and require a higher level of rigor. Evaluation engineers followed a documentation checklist developed for this track with a specific emphasis on energy model methodology and inputs. This included the following:

- Check of sources and/or reasonableness for all model inputs.
- If default inputs were used, check for reasonableness of inputs.
- Check if model files provided allow the evaluation engineer to run the model and reproduce the documented savings.
- Check that weather files were provided if applicable.
- Check that necessary information was provided to calculate demand savings.

2.3.3 Documentation Review: SEM Track

Energy Trust and the PMC provided documentation for SEM track projects that described the savings estimation approach used for program year 2021 Savings Rate Tables (SRT) as well as the documentation and data supporting the modeling approach using the Performance Tracking Tool (PTT). The following describes the two sets of documentation and how the evaluation team treated them at this phase of the project. For this evaluation year the program applied the SRT approach to claim savings, not the PTT approach, and it is therefore important to note here that the program did not spend the usual amount of time truing up the PTT workbooks.

Savings Rate Tables (SRT)

The program documented their application of the SRTs in their "2021 Savings and Forecasting Workbook". For each site, we checked the key determinants of fuel type, building type, reporting year, and engagement factor in this workbook. For spring sites, where the reporting year ended early in 2021, we corroborated a customer's engagement factor by reviewing the program's Form 101 SEM-A (for year 1 milestones) or Form 101 SEM-B (for continuation year milestones). For winter sites, where the reporting year ended late in 2021, we corroborated a customer's engagement factor from their "Log of Actions" workbooks.

Performance Tracking Tool (PTT)

The program provided the regression models and supporting data for each sampled SEM site in the form of an Excel macro-enabled workbook called the Performance Tracking Tool (PTT).

We used these PTTs as starting points for our evaluation models. Models in the PTT are generated at the monthly level since billing data is only available at the monthly level. The PTT comprises sheets describing the timing and nature of the SEM activities as well as any incentivized non-SEM capital projects, the baseline model for each fuel, along with graphical and tabular summaries of the models' residuals, total savings, net SEM savings, and incremental SEM savings for each reporting year.

The evaluation engineer followed a checklist to review the PTT documentation. This included the following:

- Documentation included accounting of non-routine events (NRE's) or capital projects installed during baseline and reporting periods.
- Meters used in the modeling identified with meter and account numbers.
- Any weather data used was present.
- Regression models provided in a form that allows the evaluation engineer to reproduce documented savings.
- Baseline and reporting periods defined (this also defines the age of the model).

2.3.4 Site Treatment

As part of the documentation review for each site, the lead analyst and QC engineer assessed whether the information and data in the project documents, along with a phone interview, was sufficient to reliably calculate savings without a site visit. If a site visit was needed, we determined if a virtual site visit was sufficient, or if a physical site visit should be performed.

We planned to reserve physical site visits for sites requiring the highest level of rigor that could not be evaluated without going on-site. However, given the mix of sampled projects present in this evaluation, we decided that conducting all evaluations virtually allowed us to obtain the necessary data to evaluate each site at the appropriate rigor level. . Figure 9 details the decision process that we used.

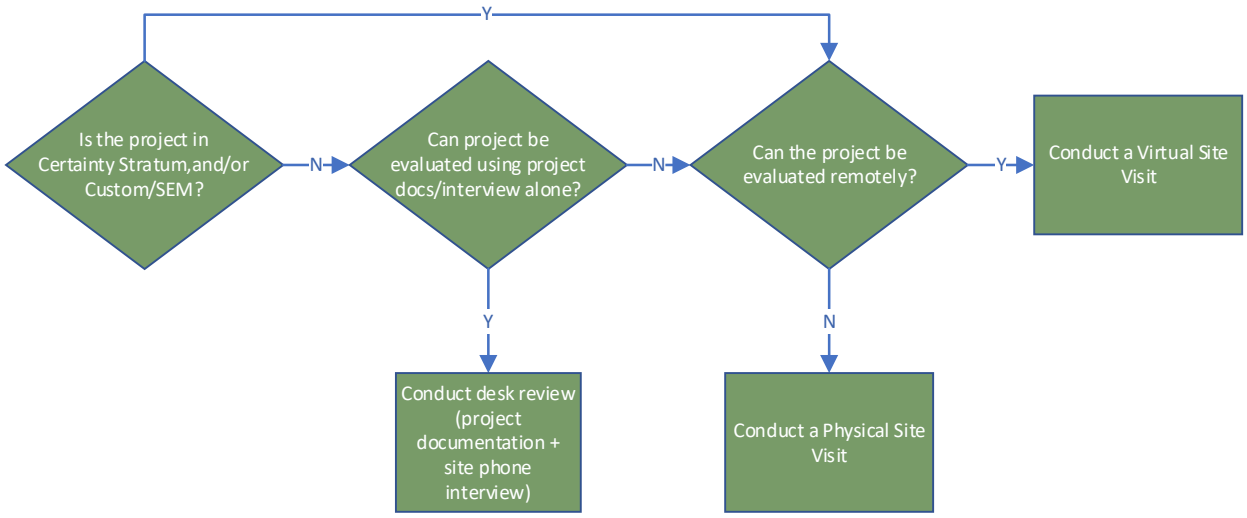

Figure 9: Determination of Site Treatment

Desk Reviews

For those sites identified as needing Desk Review only, we collected data through phone/email interviews with the site contact or other site personnel and information contained in project documentation. The information was organized for use in re-estimating savings. Special attention was placed on understanding and documenting post-installation changes in operating parameters and associated assumptions, and the implication of these changes on the estimates of energy savings. If the evaluation values were significantly different than the program values, we investigated to determine the reasons for the differences.

Virtual Site Visits

Virtual site visits included aspects of the Desk Review with the addition of photos and/or a virtual tour/interview with the customer of the installed measures. We coordinated with site personnel before the virtual visit to develop a virtual site visit agenda. A checklist was developed detailing the aspects of the site that we needed to view, and this list was shared with the site personnel prior to the virtual site visit. We anticipated that the customer would take photos and/or conduct the video tour using their own equipment/mobile devices. If devices were not available for a video tour, we explored alternatives including sending equipment on-site, or considered switching to a physical site visit.

Physical Site Visits

Note that for this evaluation year we were able to conduct all data collection with desk reviews and virtual site visits only, without the need for a physical site visit. If we had done physical site visits, they were to include all aspects of the Desk Review with the addition of an in-person visit to the site. For in-person visits, we planned to coordinate with site personnel to plan the visit. This included time for an interview and a checklist of all items to be viewed during the visit. We also planned to discuss any access requirements with site personnel beforehand, any need for

gathering trend data from the control system, and any need for on-site metering setup and retrieval. We were to then finalize the visit plan with the site contact prior to visiting.

2.4 Data Collection

This section describes the processes and procedures we used to plan and conduct data collection and interviews to support the impact evaluation. The information we gathered in this step represents new data provided by the customer beyond the project documentation we were provided.

2.4.1 Data Collection Guide

The evaluation team developed Data Collection Guides to prescribe all information needed to conduct an evaluation for all measures present at each sampled site/project such as current state of installed measures, information on facility operation, and program participation. The data collection guides are specific to each track (Lighting, Standard, Custom, SEM) as well as sections specific to each method of evaluation (desk review, virtual site visit, physical site visit). Furthermore, we created site-specific data collection guides for the Custom and SEM tracks. For all tracks, data collection guides were the template for generating a data collection workbook for each project. The evaluation analysts used these workbooks to record data gathered and results of the evaluation.

Aspects of the data collection guide for each track/site included the following:

2.4.1.1 Lighting/Standard Tracks

For Lighting and Standard tracks, the data collection guide was a collection of checklists detailing data needs by measure type/MAD. It detailed the level of rigor and list of data elements to be gathered for each measure type as well as interview questions for facility managers/staff. All sampled Lighting/Standard track projects were evaluated via desk review.

2.4.1.2 Custom/SEM Tracks

For Custom/SEM, the guides detailed elements needed to inform rigor level and collection plans specific to each project. A higher rigor level was expected for these projects and data collection included trend data, utility bills, weather data, building mechanical plans, photos of control system screens, etc. We expected most of these to require a virtual or physical site visit.

Custom Track

For Custom sites using energy models, the data collection plan included all information necessary to verify the existing energy model of the facility for both the pre- and post-installation cases. The team used this information to update the original model including building envelope, lighting systems, HVAC systems, process systems, plug loads, renewable energy generation, and any other elements needed to inform model inputs as needed.

SEM Track/Regression Models

We gathered data sufficient to generate a regression model, including baseline and reporting period weather data and meter use data for each affected fuel at the most granular level available. We focused significant data-gathering efforts on NRE detection and definition. We evaluated non-sampled capital projects at sampled SEM sites implemented during the reporting year separately, as necessary, so that savings could be accounted for appropriately.

2.4.2 On-Site Measurements/Metering

If necessary, we planned for one-time measurements to be taken on site (e.g., power readings, temperature readings, etc.) during physical site visits. We also planned to request site personnel to take measurements if reasonable. Note that for this evaluation we did not find it necessary to use our own measurements/logging for any of the sampled projects. Instead, we were able to obtain the necessary information from data supplied by the customer from control systems, trend logs, etc.

Trend logging data was obtained by capturing trend logs from customer control systems. In these cases, parameters to be analyzed were specified in the site-specific data collection plan.

2.5 Impact Evaluation Methodology

We applied the following procedures to analyze savings for each sampled measure then to use these results to estimate savings and realization rates at the program/track level.

- **Review program algorithms.** For Custom and SEM tracks, we began the analysis task by examining the program algorithms associated with the sampled measures. We did not attempt to review standardized algorithms used for the Standard/Lighting tracks as these are approved by Energy Trust separately. We took advantage of the knowledge available from Energy Trust Planning and Program and PMC staff if we had questions about savings estimation methods. In a few cases we contacted the ATAC directly after obtaining permission from Energy Trust. We documented the results of each review and reported recommendations for improvements (if any) to the algorithms that would increase the accuracy of the savings estimates. If we found algorithms that were substantially inaccurate, we noted these in the review workbook.
- **Re-estimate measure-level savings.** We adjusted algorithms for estimating savings for any substantial inaccuracies as described above. Baseline and as-built inputs to these algorithms came from our field data collection, including as appropriate, trend logging/control system information supplied by the customer and energy usage data. We analyzed pre- and post-installation utility billing data – for all sampled SEM sites and as a sanity check for calculated savings where deemed appropriate for some non-SEM sampled sites.
- **Track reasons for differences.** For all evaluated measures, to the extent possible, we documented reasons for differences between the program and evaluation savings estimates.

- **Determine COVID-19 Pandemic implications.** For projects where we identified impacts from the pandemic, we worked with the site personnel to understand their estimate for the change in values during the 2021 evaluation year as compared to the pre-pandemic period. If the site personnel indicated that pandemic impacts would remain in effect for more than half the measure life, we used the current post-installation conditions to calculate ex post savings. This approach is shown graphically in Figure 10 below.

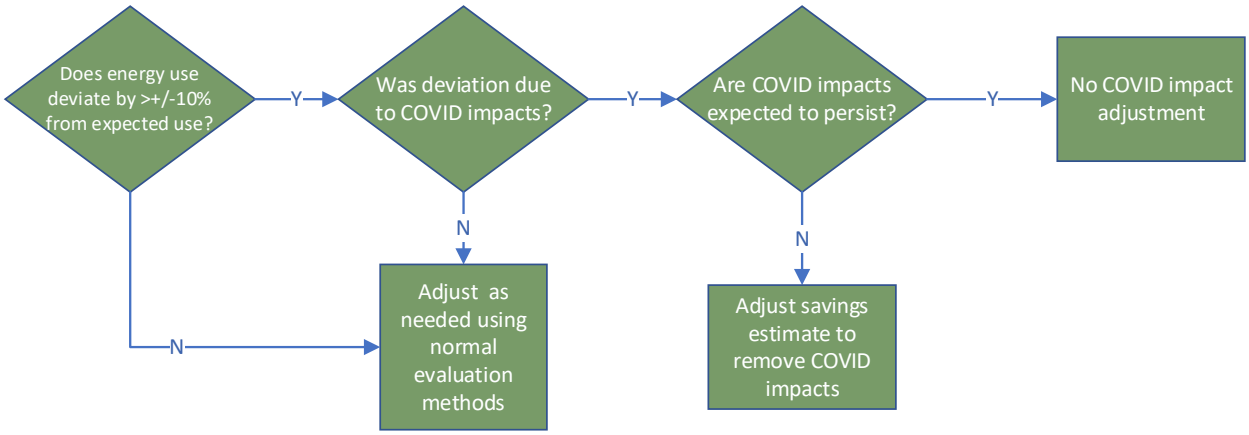

Figure 10: Determination of COVID Savings Impacts

The following discusses track-specific aspects of the impact evaluation stage.

2.5.1 Impact Evaluation: Standard/Lighting Tracks

For prescriptive/semi-prescriptive measures, the evaluation engineer examined only the inputs used to determine measure savings. The evaluation engineer did not attempt to review the MAD or savings calculators themselves but documented any recommendations for improvement observed during the evaluation. The evaluation engineer checked the measure against savings values and measure requirements from the MAD to ensure accuracy and eligibility. The evaluation engineer adjusted inputs as necessary to provide an accurate representation of savings based on data collected for the evaluation.

2.5.2 Impact Evaluation: Custom Track

Custom project evaluation included a full review of the methodology and inputs used in the savings model. The evaluation engineer modified the methodology used to determine savings as needed. Additionally, when appropriate, the evaluation engineer corrected model inputs, primarily focusing on the installed measures but may have also included changes to inputs defining the affected buildings/spaces.

For measures where the program used an hourly simulation as the basis for the ex-ante and verification estimates, we considered if changes were needed to the calibration of the model to utility bills. We attempted to recalibrate to post-retrofit utility data if major changes were made

to the original model, and/or if the original calibration was deemed inadequate. Calibration was said to be inadequate if it did not adhere to the latest PMC guidance which requires models to be calibrated to within +/-10% of annual utility billing and +/-20% of monthly billing.

2.5.3 Impact Evaluation: SEM Track

Overview

For the SEM impact evaluation, we calculated results using the program method (SRT) and modeling method (PTT). We obtained our final evaluation results using the PTT approach. We decided that this approach, as it is based on actual modeling using site-level conditions, would most likely better estimate the true site-level savings.

SRT Approach

The program calculated site-level incremental savings using the SRT approach. This semi-prescriptive approach multiplies the baseline energy usage by the program's historical saving rate (SR%) for the site's fuel type (gas, electric), building type (Higher Education, Hospital, K-12 School, MOB, Office, Parking, Prison, Public Service, and Other) and reporting year (1-5). This calculation is then further multiplied by a nominal engagement factor (EF: 0.75 – 1.25) to award more or less credit according to the customer's level of engagement over the course of the reporting year. The EF was assessed based on the number of qualifying activities that the customer carried out across their whole portfolio of enrolled building sites. Final claimed incremental savings were thus calculated at the site level as SR% x EF x baseline energy usage.

To evaluate savings using the SRT approach, we obtained the key determinants from the documentation described in section 2.3.3. We interviewed the customer's SEM energy champion at each site to better understand all SEM and non-SEM energy-impacting events and activities as well as potential non-routine events (PNREs – including COVID-19). We then updated the SRT estimate if necessary based on the data that we gathered.

PTT Approach

Prior to the COVID-19 pandemic, the program used the PTT approach to calculate savings. Models in the PTT were generated using monthly utility billing data. Although the program did not use this approach in 2021 (due to the ongoing impacts of the pandemic) to claim savings, they did generate PTTs for each site, and we used these PTTs as the basis for our site-level modeling, making adjustments as necessary.[6]

The PTT includes sheets describing the timing and nature of the SEM activities as well as any incentivized non-SEM capital projects, the baseline model for each fuel, along with graphical and tabular summaries of the model's residuals, total savings, net SEM savings, and incremental SEM savings for each reporting year.

[6] We note here that we did not attempt to compare the program-supplied PTTs to the final PTTs we used in the evaluation. We understand that the PTTs supplied to us were not as fully developed as they would have been if the PTT method were being used to claim savings.

For sites in their second, third or fourth year of SEM engagement, we considered the savings shown in the PTT for all years prior to our evaluation year to be outside our scope of work and we "locked" those savings before we made any adjustments or modifications to the PTT.

For each site we then checked the PTT for the following, listed here and described in further detail below.

- Appropriate baseline and reporting year start and end dates
- Adjustments for incentivized non-SEM capital projects implemented during the baseline or reporting periods
- Optimization of degree day base temperatures
- Removal of baseline model point outliers
- Model goodness-of-fit
- Adjustments for non-routine events

Appropriate Baseline and Reporting Year Start and End Dates

We checked to ensure that the baseline period included the correct number of months, unless it was clear from the documentation that an odd number of months were appropriate.

We also checked that each reporting year covered exactly 12 months and that the data corresponding to those 12 months were inserted into the PTT.

Adjustments for Non-SEM Incentivized Capital Projects

We reviewed the PTT entries for non-SEM incentivized capital projects to ensure they were accounted for in a manner consistent with SEM M&V guidance. We also checked that all site-specific non-SEM capital projects identified in Energy Trust's program database were entered into the appropriate PTTs with their respective savings rate adjustment factor (SRAF).

Where a non-SEM incentivized capital project was completed during the evaluation year, we obtained the associated engineering reports and calculation workbooks so that we could evaluate it separately. For those sites that we were able to evaluate, we changed the capital savings in the PTT workbook to our evaluated result and we did not apply the SRAF. Unless we sampled a non-SEM capital project under another domain, any changes we made to the capital project savings under the SEM domain did not impact the realization rate for the other sampled domains.

Optimization of Degree Day Base Temperatures

We ran the PTT's base temperature optimization routine to ensure the lowest modeling error.

Removal of Baseline Model Point Outliers

After optimizing the base temperatures, we generated a pre-post model using the daily average data for the independent and dependent variables found in the PTT's model aggregation sheet, adding a binary indicator variable to represent each reporting year. We then applied

Chauvenet's Criterion[7] to the standardized t-scores of the residuals of this pre-post model. If a t-score fell so far from zero that the probability was smaller than Chauvenet's Criterion we tagged the associated point as an outlier and removed it from the baseline model in the PTT. After omitting a point from the PTT baseline model we re-optimized the degree day base temperatures.

Model Goodness-of-Fit

We then checked the goodness-of-fit (GOF) statistics for the PTT baseline model to ensure that they satisfied the statistical criteria outlined by the SEM M&V guide. Where models did not meet these criteria we double-checked the completeness and accuracy of the raw data streams of independent and dependent variables. If the model still failed the GOF, we then checked the GOF of our aforementioned pre-post model as well as the GOF of our own forecast model which we generated using the ECAM[8] modeling application. If all three models (PTT forecast, ECAM forecast, and pre-post) failed then we used the model that had the best GOF.

Adjustments for Non-Routine Events

We qualitatively identified Potential Non-Routine Events (PNREs) through visual examination of the CUSUM[9] and residuals charts in the PTT as well as heat maps of the residuals in our pre-post model. We quantitatively identified PNREs by applying Chauvenet's Criterion to the individual t-scores of the residuals of our pre-post model (as described above for outlier detection) as well as to the t-scores for the rolling two, three, four, five, and six month averages of the individual t-scores. We then discussed these PNREs with the customer contact.

We assessed a Non-Routine Adjustment (NRA) by adding a binary indicator variable to represent the Non-Routine Event (NRE) in the aforementioned pre-post model. A negative regression coefficient for this indicator variable represents the estimated daily decrease in energy usage over the timeframe of the respective NRE, and a positive regression coefficient represents the estimated daily increase in energy usage.

We applied the NRA as an "Additional Adjustment" to our model of net savings but only if the t-statistic of the NRE indicator variable was significant *and* any one of the following three circumstances were true:

- We identified the NRE using Chauvenet's Criterion applied to the t-scores as described above

[7] Chauvenet's Criterion states that for a sample of 'n' points in a normal distribution the probability associated with any given point's distance away from the sample mean can be expected to be greater than 1/2n. If a point falls so far away from the mean that its probability is less than 1/2n then that point can be considered an outlier. This test contrasts with the typical fixed standard deviation test since it accounts for the size of the sample. Chauvenet's Criterion is discussed in IPMVP's 2020 Application Guide on Non-Routine Events & Adjustments.

[8] Energy Charting and Metrics (ECAM) is an Excel-based tool for finding and measuring energy savings. It was developed by Bill Koran and is supported by SBW Consulting, Inc. https://sbwconsulting.com/ecam/

[9] CUSUM is an abbreviation for the Cumulative SUM. It is a time-series chart of the accumulated savings over the reporting period. A change in the slope of the CUSUM reflects a change in the rate of savings.

- We identified the NRE in the CUSUM chart of the PTT as a clear and persistent change in slope starting at the onset of the COVID-19 pandemic
- The customer confirmed the NRE

2.5.4 Evaluation of Demand Savings

We calculated the electric and gas demand savings for all non-SEM sampled measures (Custom track, Lighting track, Standard track).

For electric measures using non-hourly calculations and for all gas measures, we calculated evaluation demand savings by multiplying measure-level savings by peak coincidence factors (based on measure load shape) supplied by Energy Trust.

For electric measures using hourly calculations, we used Energy Trust supplied peak periods to calculate demand savings for the measure.

As part of our demand savings evaluation, we examined the load profile used in the savings estimate. If we did not agree with the profile, we selected a more appropriate profile based on peak demand information/definitions provided by Energy Trust. In these cases, modifications made to the load profile are not included in the energy demand realization rate, but are instead reported separately. We took this approach to avoid penalizing the program for load profiles that are specific to measures (as defined in the MADs) and cannot be changed to fit site characteristics.

2.6 Population Estimation of Evaluation Results

The purpose of this task was to estimate population savings and relative precision from the sampled projects for each domain and by SEM and non-SEM program domains by fuel. We did this using stratified mean estimation.[10] For the stratified mean savings estimation method, the basic steps are as follows:

1. Mean savings for each stratum in each domain, claimed and evaluated, uses Equation 1.

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \quad (1)$$

where:

W_h $= \frac{N_h}{N}$ which is the stratum weight

N_h = population of stratum h

N = population of Program Domain

$\bar{y}_h$ = the mean of y for stratum h

[10] Cochran, William G. (1977). Sampling Techniques. New York: John Wiley & Sons.

$\bar{y}_{st}$ = the mean resulting from a stratified random sample (*st* for *stratified*).

2. Total savings for each domain for each fuel type uses Equation 2.

$$\hat{Y}_{Program\ Domain\ i} = N \times \bar{y}_{st} \quad (2)$$

3. Variance of the reporting unit mean, s^2, uses Equation 3.

$$s^2(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2\, s_h^2}{n_h} - \sum_{h=1}^{L} \frac{W_h s_h^2}{N} \quad (3)$$

where:

s_h^2 = the stratum variance and n_h is the stratum sample size.

The second term in the equation represents the finite population correction.

4. Relative precision of the reporting mean at 90% confidence uses Equation 4.

$$RP_{Program\ Domain\ i} = \frac{1.645 \times s(\bar{y}_{st})}{\bar{y}_{st}} \quad (4)$$

where:

$s(\bar{y}_{st})$ = the standard error of the stratified mean

Then we calculated Realization Rate (RR) for each domain by fuel and across the SEM program domains and non-SEM program domains by fuel as the ratio of the population estimate of evaluated savings to total claimed savings for each level, as shown in Equation 5.

$$RR_{Summary\ Level} = \frac{\text{Evaluated } \hat{Y}_{Summary\ Level}}{\text{Claimed } \hat{Y}_{Summary\ Level}} \quad (5)$$

We applied Equations 6 through 8 to calculate the relative precision of the SEM, non-SEM, and overall estimates.

$$EB_{Program\ Domain\ i} = RP_{Program\ Domain\ i} \times \hat{Y}_{Program\ Domain\ i} \quad (6)$$

For estimating the RP for the realization rate at the higher summary levels, we propagated the Program Domain error bounds (EB Program Domain) across each sublevel to the next higher level using Equation 7, with RP calculated as Equation 8.

$$EB_{Summary\ Level} = \sqrt{\sum (EB_{Program\ Domain\ i})^2} \quad (7)$$

$$RP_{Summary\ Level} = \frac{EB_{Summary\ Level}}{\sum \text{Evaluated } \hat{Y}_{Program\ Domain\ i}} \quad (8)$$

3 Findings

In this section we detail all reporting of results and recommendations to Energy Trust.

3.1 Final Sample Disposition

Table 6 shows the final sample disposition. Total number of evaluated projects and measures are shown, along with the percent of projects completed relative to the original sample.

Note that for the Lighting and Custom tracks, we did not complete 100% of the original sample. We experienced a number of challenges related to recruitment that restricted the number of sites we were able to complete. This affected the final relative precision for the sample as well. Specific recruitment challenges and the methods we used to address them are discussed directly below.

In the Custom track, we had one site in the electric and one site in the gas certainty stratum that were not evaluated due to lack of response from the customer. To account for this, we moved these sites from the certainty strata (strata 9) to the next largest sampled strata (strata 3). This affected the relative precision around the Realization Rate estimate minimally for the electric frame (increased by about 1%) and negligibly for the gas frame.

For the Standard and SEM tracks, we were able to complete more evaluations than originally sampled.

Table 6: Sample Disposition

Program Track	Electric			Gas		
	Completed Sample		% Complete	Completed Sample		% Complete
	Projects	Measures	Projects	Projects	Measures	Projects
Standard	31	39	103%	26	31	104%
Custom	10	13	43%	15	17	75%
Lighting	25	129	83%			
Capital	**66**	**181**	**80%**	**41**	**48**	**91%**
SEM	**27**	**27**	**108%**	**26**	**26**	**104%**
Grand Total	**93**	**208**	**86%**	**67**	**74**	**96%**

3.1.1 Recruitment Challenges

In a typical evaluation, it is normal for evaluators to experience some challenge in recruiting customers. There can be various reasons for this, including customers going out of business, turnover in personnel, customer refusal to participate, etc. For the 2021 evaluation, similar to 2020, we experienced a significantly higher number of challenges and issues than in previous evaluations in both getting customers recruited and in getting customers to continue cooperation

throughout the evaluation process. Our recruitment rates per track (number of sites recruited / number of sites contacted) is shown below in Table 7 for 2021 (with 2020 included for comparison).

The overall percent recruited was 77% in 2020 and improved to 84% in 2021. Note that these recruitment rates do not reflect numerous cases where customers had been recruited but later ceased cooperation with the evaluation. For instance, we experienced a large number of Custom track customers who had been recruited at the outset but later became unresponsive and did not follow up on our requests to provide data. This resulted in a significant number of dropped projects in the Custom sample despite the higher recruitment rate.

The number of recruitment issues had the potential to greatly reduce the statistical significance of results. Due to the high number of issues, we did not use site replacement to mitigate this as replacement sites were likely to have the same recruitment issues.

Many of the challenges stemmed from the impacts of the COVID-19 pandemic. In 2020 itself, most businesses experienced some sort of impact and in many cases these impacts have extended post-pandemic into 2021 and beyond. During our recruitment process, we found that businesses experienced significant personnel reductions and/or changes. Original site contacts were either no longer working for the customer or had taken on additional roles and become too busy to participate in the evaluation. New contacts often had no knowledge of the installed measures.

To mitigate these challenges, we attempted multiple contacts by both phone and email. If a customer still proved unresponsive, we worked with the PMC to attempt re-contacts. If this further step was unsuccessful, rather than immediately dropping the project from the sample, we considered whether or not we could reasonably evaluate savings at the site using project documentation alone (a "no-contact" review). For a no-contact review, we required some sort of documented proof of measure installation in the provided documentation package such as invoices, inspection reports, etc. We also attempted to confirm that the customer was still in operation by checking for current websites or calling the business to see if phones were still being answered. We discussed this approach with Energy Trust for approval before proceeding with no-contact reviews.

Table 7 shows the 2021 recruitment rate, the 2020 recruitment rate for comparison, and the number of no-contact reviews we performed for 2021 per program track. We did not perform no-contact reviews for Custom sites due to the more complex data needs for those sites.

Table 7: Recruitment Rates and No-Contact Reviews

Track	Recruitment Rate (This Evaluation)	Recruitment Rate (2020 Evaluation)	Number of No-Contact Reviews Conducted
Standard	82%	80%	8
Custom	89%	82%	0
Lighting	59%	77%	10

Track	Recruitment Rate (This Evaluation)	Recruitment Rate (2020 Evaluation)	Number of No-Contact Reviews Conducted
SEM	98%	73%	1
Overall	**84%**	**77%**	**19**

Despite recruitment challenges, we were able to complete a sufficient number of project evaluations to result in a reasonable relative precision overall. Our original goal was a 90/10 sample (a relative precision of 10% at a 90% confidence level) for each fuel by SEM and non-SEM projects. Final relative precisions are shown in Table 8 and Table 9 below.

3.2 Evaluation Savings and Realization Rates

In this section we discuss the overall evaluation savings and resulting realization rates. For this discussion, measure domains are grouped by Capital (all non-SEM projects) and SEM. The Capital grouping includes the Standard, Lighting, and Custom program tracks. For the SEM track, both SRT- and PTT-based results are shown in Table 8 and Table 9, and are discussed below, but our evaluation findings are based on the PTT results. As PTT was used as our final result, it is the primary focus of the results discussions that follow. We do however discuss the SRT approach as well and how it might be improved upon so that results align better with realized (modeled) savings.

Note that the relative precision in the final results, although affected by completed project/savings magnitude, is also affected by the scatter of realization rates within each domain. Domains with large scatter tended to have less precision. For example, the SEM gas and electric samples had a number of both high and low realization rates, and this resulted in a higher uncertainty in the program-level results.

Measure level results for all individual measures completed in the sample are listed in Table 14 in the Appendix.

3.2.1 Electric Energy Savings and Realization Rates

Table 8 shows electric energy savings and realization rates (RR) by domain, along with the relative precision at 90% confidence for each result. Across all tracks, electric savings showed a realization rate (93%), with most domains near this number. The most notable exceptions were a high RR for the Standard track domain, and low RR for the Custom domain.

Note that the precision for both the Custom and SEM results were low compared to the original precision targets. The low precision for Custom was the result of a low customer response rate for this track and hence a lower number of evaluated projects. For this reason, we consider this result uncertain.

The low precision for SEM was the result of a high degree of scatter between claimed (SRT-based) savings and evaluated (PTT-based) savings. This is an important observation as the low precision is not reflective of the PTT method itself but is more a reflection of the accuracy in the savings estimates between the PTT and SRT approaches. Although the SRT method yielded a higher precision in the realization rate, comparison with the PTT approach at the project level indicates that the PTT method is the more accurate approach to calculating project-level savings.

Table 8: 2021 Electric Energy Savings and Realization Rates by Domain

Program Track	Electric Frame							
	Unique Measure Count	Unique Project Count	Evaluated Project Count	Population Claimed Savings (kWh)	% Savings Sampled	Population Evaluated Savings (kWh)	Realization Rate	Relative Precision @ 90% Confidence
Standard	405	318	31	16,858,042	41	17,848,238	1.06	13.3
Custom	247	177	10	26,913,894	10	17,737,402	0.66	34.5
Lighting	3,886	1,149	25	54,948,182	13	56,060,433	1.02	10.0
Capital	**4,538**	**1,644**	**66**	**98,720,118**	**17**	**91,646,073**	**0.93**	**9.4**
SEM - SRT	**449**	**449**	**27**	**12,701,358**	**32**	**10,634,142**	**0.84**	**15.7**
SEM - PTT	**449**	**449**	**27**	**12,701,358**	**32**	**11,492,829**	**0.90**	**67.9**
Grand Total w/ SEM SRT	**4,987**	**2,093**	**93**	**111,421,476**	**19**	**102,280,215**	**0.92**	**8.6**
Grand Total w/ SEM PTT	**4,987**	**2,093**	**93**	**111,421,476**	**19**	**103,138,902**	**0.93**	**11.3**

3.2.2 Gas Energy Savings and Realization Rates

Table 9 shows gas energy savings and realization rates by domain, along with the relative precision at 90% confidence for each result. Gas savings showed a slightly lower realization rate overall than electric at 89%, with most domain RRs near 1. The only domain falling below a 90% realization rate was the SEM domain (PTT-based). Track-specific results are discussed in more detail below.

As with the electric SEM and Custom results, the causes for low precision were lower number of completed project evaluations for Custom, and large scatter in site-level realization rates for SEM.

Table 9: 2021 Gas Energy Savings and Realization Rates by Domain

Program Track	Gas Frame							
	Unique Measure Count	Unique Project Count	Evaluated Project Count	Population Claimed Savings (Therms)	% Savings Sampled	Population Evaluated Savings (Therms)	Realization Rate	Relative Precision @ 90% Confidence
Standard	172	140	26	968,411	39	925,691	0.96	11.8
Custom	149	114	15	1,021,437	24	1,007,478	0.99	25.5
Capital	**321**	**254**	**41**	**1,989,848**	**31**	**1,933,169**	**0.97**	**14.4**
SEM - SRT	**379**	**379**	**26**	**481,830**	**30**	**471,415**	**0.98**	**10.2**
SEM - PTT	**379**	**379**	**26**	**481,830**	**30**	**267,565**	**0.56**	**85.1**
Grand Total w/ SEM SRT	**700**	**633**	**67**	**2,471,678**	**31**	**2,404,585**	**0.97**	**11.8**
Grand Total w/ SEM PTT	**700**	**633**	**67**	**2,471,678**	**31**	**2,200,734**	**0.89**	**16.4**

The following sections discuss results by program track.

We have provided "waterfall" charts in these sections for non-SEM tracks to graphically illustrate estimated changes in measure-level energy savings by reason. Claimed savings are shown as a bar on the left of the chart, with final evaluated savings shown as a bar on the right end. In between the two, primary reasons for differences are listed, with a qualitative magnitude of impact on savings (based on engineering judgment) shown. "Cases" in these charts represents the estimated primary reason for change in savings for each adjusted measure. Note that in instances where a measure had multiple reasons for savings changes, secondary reasons were omitted from these charts for brevity. Cases with "no change" represent the number of measures in the domain that did not change.

3.2.3 Standard Track Results

Standard track documentation was found to be excellent, with all items necessary to perform an evaluation present for the vast majority of sampled projects. This included application, savings worksheet, invoices, equipment cutsheets, and any applicable measure exceptions.

We were unable to recruit 10 projects in the Standard track due to lack of response from the customer, but documentation was sufficient that we were able to complete no-contact reviews for 9.

Overall, the kWh realization rate for the Standard track was 1.06 with 31 projects evaluated (1 more than the original sample of 30). The higher RR for this track was the result of three projects with high realization rates. Two of these projects (RRs of 6.3 and 12.3) specified

installation of gas measures but installed measures were electric, and one project (RR of 4.8) was missing an installed measure in the savings claim. The estimated effects of realization rate by reason can be seen in Figure 11 below. Note that the three projects with high realization rates are listed under the "Other" reason category, along with one other measure (reason for difference was incorrect measure application used). Most projects received a realization rate at or near 1.0. We did not find any projects with zero savings.

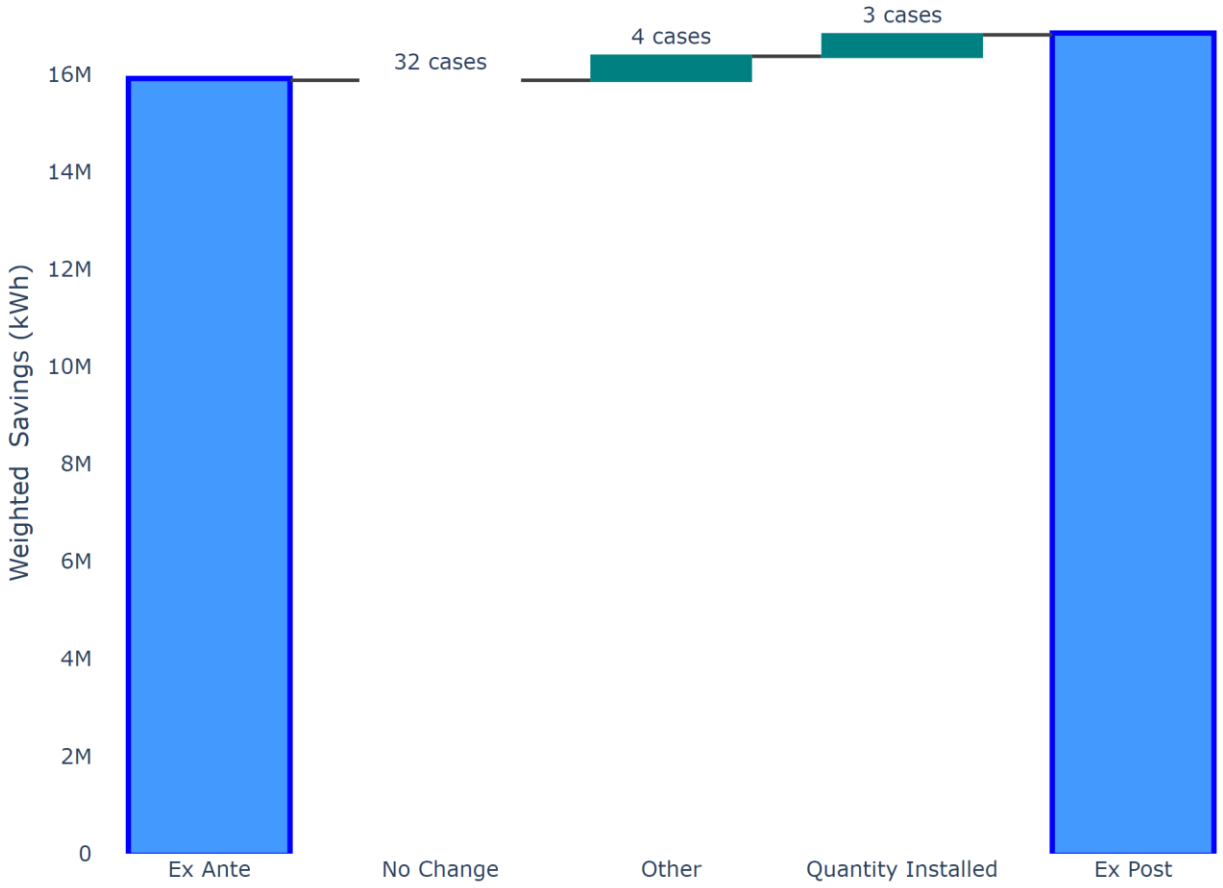

Figure 11: Decomposition of Electric Energy Use Change by Reason (Standard Track)

The therm savings realization rate for Standard Track was 0.96, representing 26 evaluated projects (1 more than the original sample of 25), of which 4 received no-contact reviews. Most projects in this portion of the sample received RRs of close to 1.0. We found two projects with zero savings. In one case, the customer had gone out of business and the measure was no longer operating (the building was unoccupied/off-line with plans to completely electrify it in the future). In the other case, also mentioned above in the electric discussion, the installed measure was electric rather than gas as specified in the original savings claim.

Primary reasons for differences are shown in Figure 12. Reasons in the 'Other' category included: number of measures installed, incorrect fuel used for measure, incorrect steam capacity used, incorrect building type used, and wrong measure application used.

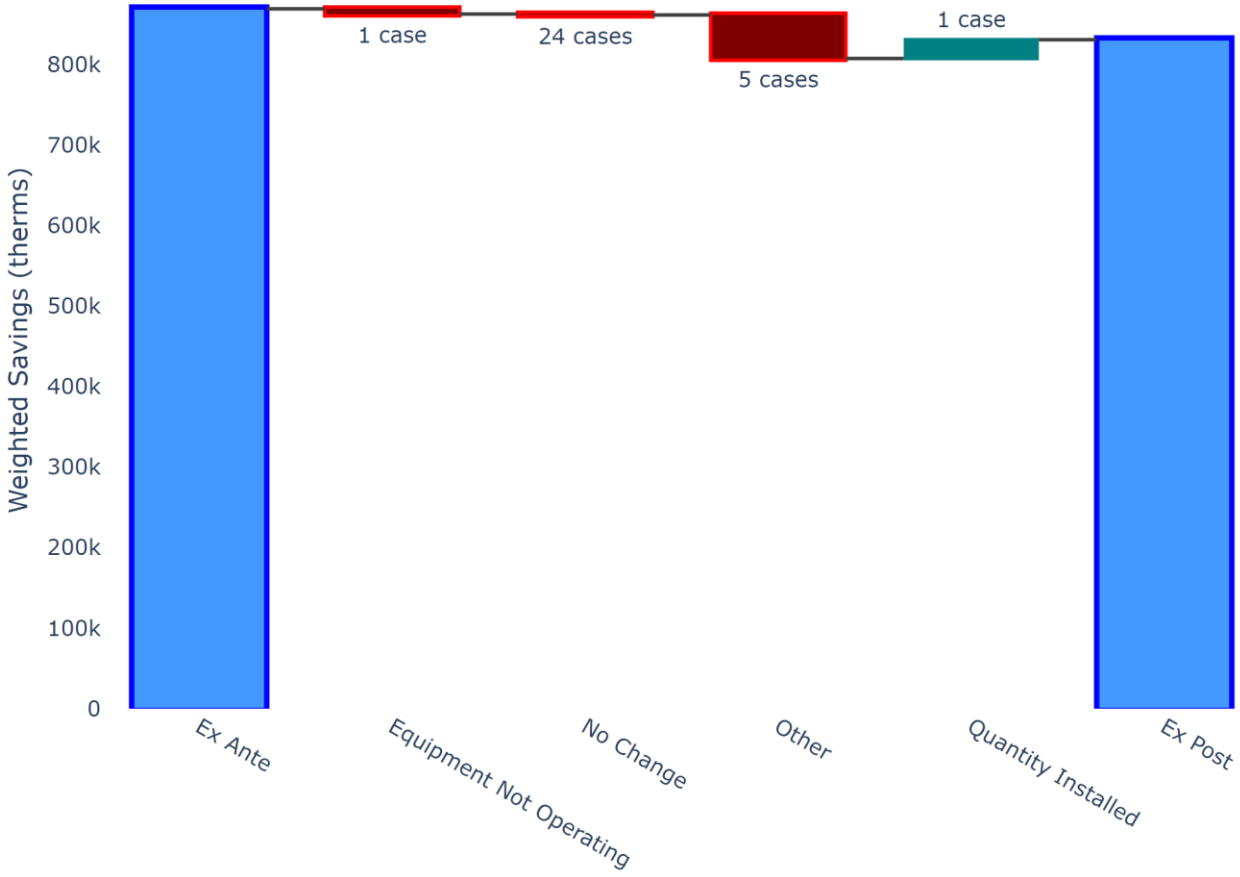

Figure 12: Decomposition of Gas Energy Use Change by Reason (Standard Track)

3.2.4 Lighting Track Results

In this section we detail evaluation results for the Lighting track (kWh fuel only).

Lighting projects were completed under a variety of MADs and use of the Energy Trust Lighting Tool. We found documentation to be adequate in describing proposed measures, with some post-installation documentation in the form of post-inspection reports, but many times itemized invoices for the installed measures were not available. We were able to evaluate 25 of 30 originally sampled projects. We were unable to recruit 15 customers, but were able to perform no-contact reviews for 10 sites. The remaining 5 unresponsive sites were dropped from the sample due to lack of adequate post-installation documentation.

We calculated the realization rate for the Standard Lighting Domain to be 1.02. We found 3 measures in which product wattage was different than proposed. In two cases, measures were found to be ineligible because the specified equipment was not listed in the required QPL. For 6 projects we found quantity of fixtures installed to be different than proposed. Reasons for differences are shown in Figure 13 which shows graphically the effects of the primary changes in electric energy savings. Note that secondary changes are excluded from this chart for brevity.

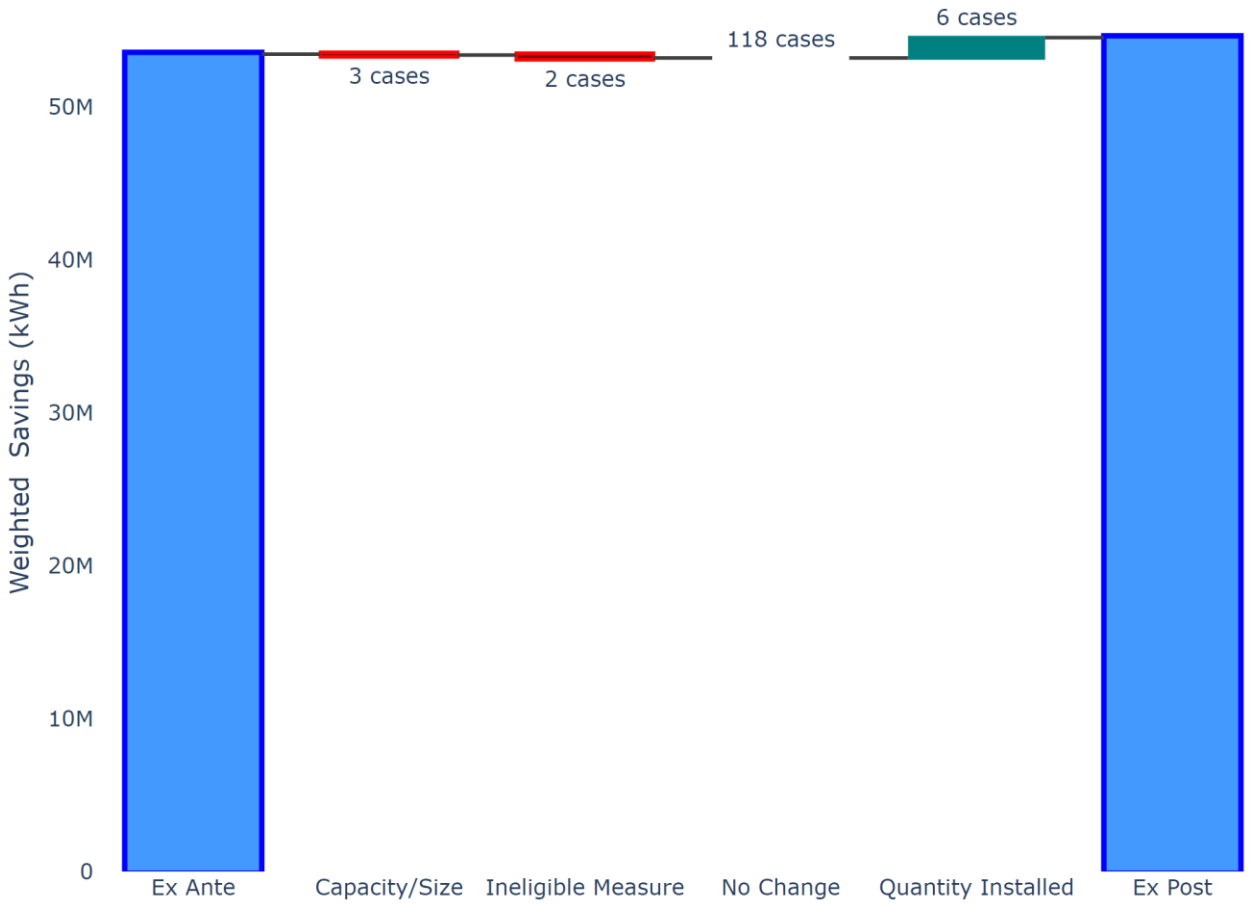

Figure 13: Decomposition of Electric Energy Use Change by Reason (Lighting)

3.2.5 Custom Track Results

We completed 10 of 23 originally sampled electric project evaluations and 15 of 20 sampled gas evaluations. Due to the complex nature of the remaining unresponsive sites, we were not able to perform no-contact reviews and were forced to drop them from the sample. The relative precisions in this domain are therefore lower, indicating uncertainty in these results (see discussion of final sample disposition in 3.1 above).

Figure 14 and Figure 15 show graphically the effects of the primary changes in electric energy and gas savings respectively. Note that secondary changes are excluded from these charts for brevity. Reasons for changes varied widely between and among projects.

We found the electric realization rate to be 0.65. On the electric side, the largest estimated change in savings was due to issues involving setpoints used in the model. These required various adjustments to reflect setpoint values from current site conditions. Other reasons varied widely as can be seen in Figure 14.

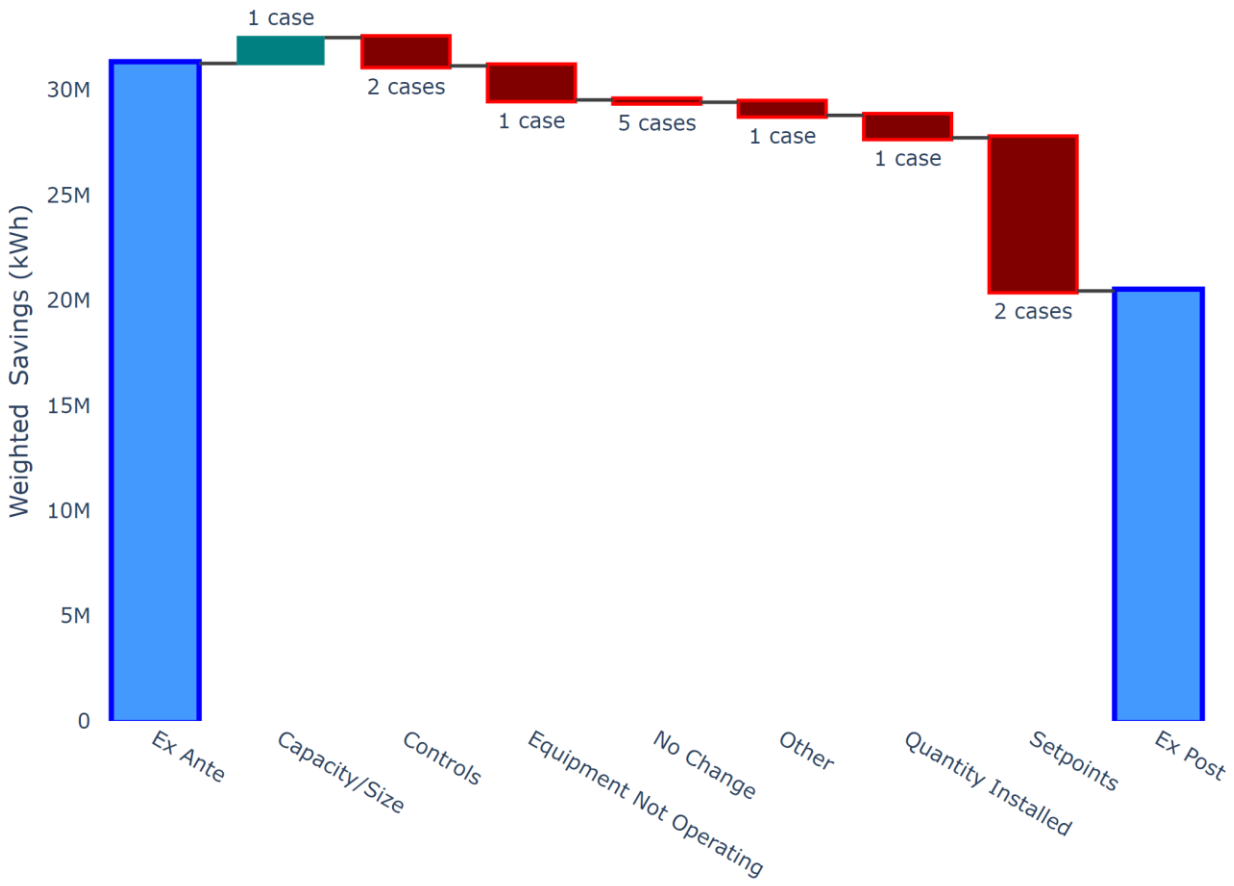

Figure 14: Decomposition of Electric Energy Use Change by Reason (Custom Track)

The gas side realization rate was higher than the electric side at 0.99. The largest changes in savings included two cases of incorrect baseline inputs (causing savings to increase overall) and one case of equipment found not to be operating (causing savings to decrease overall). Other reasons caused minor differences in overall savings.

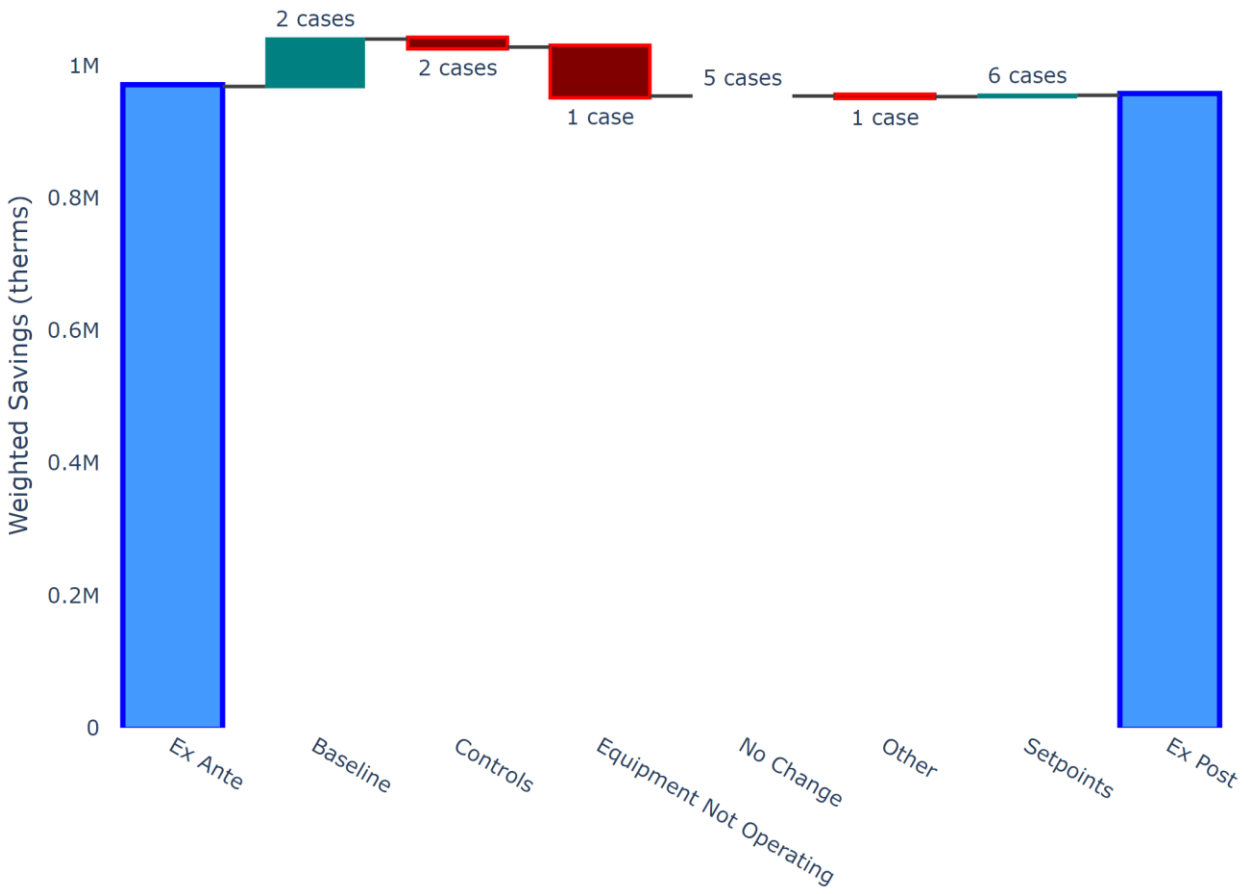

Figure 15: Decomposition of Gas Energy Use Change by Reason (Custom Track)

As part of the Custom track evaluation, we also checked model calibration to utility billing for each model that used an 8,760 hour whole building simulation. Of the 16 models checked, we found 12 to have some sort of calibration issue, meaning that the calibration was inadequate compared to the latest PMC guidance. The latest guidance requires models to be calibrated to within +/-10% of annual utility billing and +/-20% of monthly billing. Of the 12 models with calibration issues, we completed evaluations for 6 (for the remaining 6 we did not complete an evaluation due to lack of customer response). Five completed sites had calibration issues such that we were not able to re-calibrate the model. We were able to calibrate one model to within 5% of annual billing data, with a reasonable fit to monthly usage, but some months were outside the 20% prescribed by the PMC guidance. Calibration issues for the six sites included the following:

- Utility billing data for the sampled fuel was not available.
- Utility meters spanned multiple buildings so that data for the target building could not be extricated.
- The model left out multiple systems or areas covered by the utility meter to the extent that a major model re-build and re-calibration would be necessary.

For some of these models, we were able to complete a utility billing check to ensure that apparent savings between the pre- and post-retrofit period was reasonable compared to evaluated savings. For others, we were not able to perform a utility bill check because either utility data was not available, or utility meters spanned multiple buildings and data for the target buildings could not be extricated. For all six sites, we accepted savings after completing a full model review and adjusting inputs as necessary based on information gathered from the customer.

We also note here for context that the majority of sampled Custom Track projects were implemented by the previous PMC, so many of the issues found for this track were not related to projects implemented by the current PMC.

3.2.6 SEM Track Results

In this section we discuss results for the SEM track. We first briefly describe our findings for the SRT and PTT approaches, followed by a brief discussion comparing the savings estimated with the SRT to the savings calculated with the PTT. We obtained our final evaluation results using the PTT approach, with adjustments to account for COVID-19 impacts.

Table 10: SRT- and PTT-based Realization Rates

	SRT Approach		PTT Approach	
	Realization Rate	RP @ 90% Confidence	Realization Rate	RP @ 90% Confidence
Gas	0.98	10	0.56	85
Electric	0.84	16	0.90	68

SRT Results

Following the SRT approach, we assessed the electric and gas savings to be 84% and 98% of what the program claimed, respectively.

We found discrepancies with the program's claimed savings at seven sites. One was due to the building being dropped from the program; one was due to the PTT model age of four years old (instead of three years old[11]); and the other five were due to discrepancies we found in the number of qualifying projects that were counted toward assessing the engagement factor. Those five were all spring sites for which the "Log of Actions" documents were not yet in circulation at the end of the reporting year as they were for the winter sites.

PTT Results

To calculate evaluation savings for the PTT approach, we applied adjustments where we found COVID-19 impacts to be significant. We reported zero savings if our evaluated incremental

[11] The historical savings rate for four year old models is less than the savings rate for three year old models.

savings were negative. Using the PTT approach, the relative precision (uncertainty) of both the gas and electric results (see Table 10) at the population level is high due to the extreme variance in the realization rates across the sample.

Electric Savings

We evaluated the electric savings to be 90% of what was claimed, compared to 84% for the SRT approach. We found 13 of the 25 electrical sampled sites to have negative incremental savings, which we reported as zero savings. Sampled site-level results are provided in the appendices.

Gas Savings

We evaluated the total gas savings to be 56% of what was claimed, compared to 98% for the SRT approach. We found 16 of the 24 gas samples to have negative incremental savings, which we reported as zero savings. Sampled site-level results are provided in the appendices.

COVID-19 Impacts

We found significant COVID-19-related NREs in 16 of 25 electric sampled sites and 6 of 24 gas sampled sites. We assessed the incremental savings before and after applying the COVID-19 NRAs.

When we applied the COVID-19 NRAs, savings became positive where they had been negative at three of the electric samples and one gas sample, so COVID-19 essentially had the effect of negating savings at those sites. While we could not confirm for certain, those impacts may have been due to SEM backsliding.

Conversely, savings became negative where they had been positive at eight electric samples and two gas samples. In other words, operational changes related to the COVID-19 pandemic had the effect of creating savings at those sites. We zeroed the negative savings at those sites.

PTT Adjustments

The following summarizes the adjustments we made to the PTTs for evaluating savings using the PTT approach.

- **Baseline and Reporting Year Start and End Dates** We adjusted the baseline and/or reporting year timeframes in nine PTTs to ensure they covered exactly 12 months. However, at one site we reduced the baseline timeframe from 36 months to only 11 months to avoid the bias and noise associated with an earlier renovation project.
- **Adjustments for incentivized non-SEM capital projects implemented during the baseline or reporting periods** We inserted non-SEM incentivized capital projects in eight PTTs where they had been omitted. We adjusted the claimed savings for one capital project based on our own findings. At one site we adjusted the completion date of a capital project.
- **Removal of baseline model point outliers** We removed baseline point outliers in eight models.

- **Optimization of degree day base temperatures** We optimized the degree day base temperature in 39 of the 48 PTTs.
- **Model goodness-of-fit (GOF)** We found GOF to be out of compliance in three models. One of those we brought into compliance by shifting the baseline timeframe back four months, and another by omitting a point outlier from the model. The third one was at the site where we reduced the baseline timeframe from 36 months to 11 months. While reducing the baseline timeframe improved the model, it still did not bring it into compliance. For that site we finally found the best fit in the form of a pre-post model, albeit still out of compliance.
- **Adjustments for non-routine events** We found significant COVID-19-related NREs at 21 sites and other NREs at seven sites. Three of the 21 sites impacted by COVID-19 were also impacted by other NREs.
- **Other issues**
  - **Failed Meter** We found one site where the data for the independent variable was from a meter that had failed over the last four months of the evaluation year. For that site we used ECAM to generate a model of the 12 months preceding the meter failure and used that model to predict what the meter readings would have been over the last four months of the evaluation year. With those four months of data filled in, we then used ECAM to calculate the avoided energy for the actual 12-month evaluation year.
  - **Administrative Error** One site had no savings because it had been dropped from the program during the evaluation year.
  - **FSU > 100%** We found nine sites where the uncertainty in the total modeled savings for the year (at the 68% Confidence Level) was greater than the total modeled savings. Similarly, we found seven sites where the uncertainty in the total modeled savings was greater than the incremental savings.

Comparison of SRT and PTT Results

At the population level, the SRT approach agreed closely with the PTT approach for electric but not for gas. Using the PTT approach, we evaluated the total electric savings to be 107% (0.90/0.84) of what we assessed using the SRT approach, and the total gas savings to be 57% (0.56/0.98) of what we assessed using the SRT approach. This means that the SRT approach as we applied it appears to have underestimated the electric savings by 8% but overestimated the gas savings by 43%. However, given the poor relative precision in the PTT results at the population level, we cannot draw any general conclusions from this comparison. For example, on the gas side, the main reason for the low overall realization rate was that the PTT approach returned zero savings for the same five samples where savings were highest by the SRT approach.

3.2.7 Demand Savings Results

Summer and Winter peak demand savings were calculated for all tracks except SEM. Table 11 And Table 12 show demand realization rates by track, with energy consumption realizations rates shown for comparison.

Table 11: Comparison of Electric Energy and Demand Savings Realization Rates by Track

Program Track	kWh RR	Summer kW RR	Winter kW RR
Standard	1.06	1.10	1.00
Custom	0.66	0.77	0.84
Lighting	1.02	1.00	1.00
Overall	**0.93**	**0.98**	**0.99**

Table 12: Comparison of Gas Energy and Demand Savings Realization Rates by Track

Program Track	Gas Consumption RR	Gas Demand RR
Standard	0.96	0.97
Custom	0.99	0.39
Lighting	NA	NA
Overall	**0.89**	**0.90**

Custom had the smallest kW demand savings realization rates for both summer and winter, due mainly to changes in kWh savings during peak periods. Standard and Lighting projects had RRs higher than 1 due to increased kWh savings overall. Note that lighting kW demand RRs are in actuality greater than "1.00" when more decimal spaces are used (i.e. these values are rounded to two decimal spaces).

On the gas side, Custom also had the lowest gas demand realization rate, with Standard being near 1.

We found that some profiles for prescriptive measures (Lighting and Standard tracks) did not match the site's specific use/building type. We noted which profiles we deemed to be correct, but did not change evaluated demand savings by adjusting these profiles because the profile itself is specified in the MAD and cannot be changed. If corrected profiles were included in demand savings calculations, results would be as shown in Table 13.

Table 13: Profile-Adjusted Electric and Gas Demand Savings Realization Rates by Track (shown for information only)

Program Track	Electric Demand Summer RR	Electric Demand Winter RR	Gas Demand RR
Standard	0.85	1.48	1.01
Custom	0.77	0.90	0.39
Lighting	1.01	1.00	NA
Overall	**0.90**	**1.31**	**0.94**

3.3 Observational Questions and Previous Recommendations

In this section we present answers to Energy Trust's observational questions and provide our response to recommendations from previous recent evaluation reports.

3.3.1 Observational Questions

The following questions were asked by Energy Trust staff. We present the questions along with any findings here.

- Are there project files for every site and do those files contain complete information? Are there obvious errors in any of the assumptions used in the energy analysis?

  Findings: We found numerous issues with project documentation during the course of the evaluation. These included items missing from the document package or incorrect versions of the technical analysis study, savings calculation files, etc. We observed three categories of documentation issues overall: missing documentation, incorrect model/calculation version provided, and incorrect document version provided. We found the fewest issues with the Standard track projects. We found the highest number of issues with SEM track projects, followed by Custom. Below is a summary of our findings by documentation issue category. Note that errors in energy analysis assumptions are covered in the Energy Use Change by Reason graphs in the Findings section above for each track.

  - Missing Documentation (documentation critical to performing evaluation was missing): Custom (14 issues), Lighting (9 issues), SEM (3 issues), Standard (2 issues).
  - Incorrect Model/Calculation Version (savings from model or calculation files provided did not match claimed/documented savings): SEM (7 issues), Custom (1 issue).
  - Incorrect Document Version (documentation provided did not match claimed savings, or wrong version of documentation was provided): SEM (23 issues), Custom (2 issues), Lighting (1 issue).

- Were there any post-installation changes in operating parameters and associated assumptions? If so, what were the consequent changes in energy savings estimates for individual projects (e.g., changes in operating hours)?

  Findings: We found a variety of changes to post-installation operating parameters/assumptions. These are covered in the Energy Use Change by Reason graphs in the Findings section above for each track. Resulting changes are captured in the realization rates for each project.

- What are the factors that result in large variances in energy savings from program estimates (e.g. assumptions too conservative, incorrect hours of operation)?

  Findings: These are covered in the Energy Use Change by Reason graphs in the Findings section above for each track.

- For Custom track measures, are there trends in savings realization by ATAC firm completing the energy study?

  Findings: Projects in the Custom track were spread across 13 ATACs. We analyzed average electric and gas realization rates by ATAC and did not find a detectable pattern of realization rate by ATAC (i.e. the average ATAC realization rate did not deviate from the overall average realization rate by an amount that would suggest an issue). Also note that in our sample, no ATAC had more than 3 projects, so the sample size for determining results by ATAC is likely too small to draw any reliable conclusions.

- Are the projects using the appropriate baseline (existing conditions or current market) to estimate savings and cost-effectiveness? (Note: results may not be statistically significant but will provide an overall check).

  Findings: We did not identify any significant issues with specification of baseline type in any program track. This indicates that baseline type in general is being applied appropriately.

- For Standard track measures, do the measure approval documents used by the program include sufficient information to estimate reliable savings, and if not, what specific changes should be made to improve them?

  Findings: As discussed above, we checked each applicable measure against the appropriate MAD to ensure that the measure met all eligibility requirements and was specified correctly. During this process, we did not identify any issues with the MADs and found them to be well laid out and easy to use. Note that the research question specifically mentions the Standard program track, but this finding applies to the Lighting program track as well.

- Were recommendations made in previous impact evaluations implemented, and if so, how have these changes affected the realization or verification of program savings?

  Findings: These findings are covered in their own section (3.3.2 below).

- What recommendations does SBW have regarding analysis approaches and assumptions, or customer behavior or decision-making that would be helpful to Energy Trust in designing, implementing, and evaluating its programs in the future?

  Findings:

  - **Specific Track Recommendations:** See Evaluation Observations/Recommendations in the Executive Summary.
  - **Customer Behavior/Decision-Making:** Due to the recruitment difficulties mentioned above, we recommend emphasizing to customers the possibility and need for evaluation throughout the customer engagement process. This could include reminding customers of the requirement to participate in evaluation per their incentive contract, or offering some sort of incentive for participating. Also, if the PMC is aware of a change in the site contact, we suggest notifying the new contact of the possibility of evaluations before the evaluation process begins. For future evaluations we recommend continuing to focus on recruitment process both in terms of the actual procedure used to recruit as well as timing (begin recruitment as early as possible).

- Are there economic or other trends that are impacting the program's ability to forecast and estimate savings? (Note: We will study impacts qualitatively).

  Findings: In the 2020 evaluation we observed significant changes to the customer population (likely as a result of the COVID pandemic) that likely impacted the program's ability to forecast and estimate savings. All of these observations also apply to this evaluation (2021). These observations include:

  - Many offices and similar building types experienced drastically reduced occupancy. Although occupancy appears to have rebounded, many customers indicated they were unsure whether or not occupancy levels would return to pre-COVID levels.
  - Many customers experienced significant organizational changes as a result of the pandemic including staff downsizing, loss of personnel, reduced roles/hours, mergers/buyouts, etc. This resulted in significant challenges to us (the evaluators) in contacting and recruiting customers. These challenges persisted for the 2021 evaluation. Based on what we observed, we feel like these impacts will persist into at least the near future. These impacts may be presenting challenges to the program's interaction with their customer base.
  - We observed that many education and healthcare facilities increased their ventilation (outside airflow) to help cope with the pandemic. Many if not most of these customers indicated that they either did not plan to return to pre-pandemic ventilation levels, or they had planned to do so but plans were never implemented. We observed significant increase of energy use at these facilities due to this, and we feel it is likely to persist into at least the near-term future. While this increases energy use, it may also present additional opportunities for energy savings (better controls, O&M, etc.).

- For SEM track measures, how well did the deemed savings approach approximate actual energy savings?

  Findings: These results are reflected in the SEM realization rates, discussed in the SEM Track Results section, which compare savings from the SRT method used by the program with evaluation savings (PTT approach w/COVID impacts accounted for) as modeled by SBW.

3.3.2 Previous Report Recommendations and Status

In this section we list the recommendations from the 2020 evaluation report and provide an update on the status of each. In 2020, we had recommendations pertaining to the Lighting and SEM tracks.

Lighting track

- **Small Commercial Direct-Install Lighting Projects (MAD 18)/Standard Lighting Projects (various MADs) 2020 Report Recommendation:** "We were not able to obtain invoices for these project types (all 11 DI Lighting projects and all 11 Standard Lighting projects) from available Energy Trust documentation, and many times customers were not able to supply an invoice at our request. This made it difficult to evaluate these projects. In particular, we did not have a manufacturer/model of installed units to check against the Qualified Products List (QPL).

  Recommendation: Require the installer/customer to supply the itemized invoice showing quantity, manufacturer, and model number of installed units. Note that this may be difficult due to the nature of direct installs, but it would be very helpful in documenting the installation of the measures."

  Findings: We did not evaluate direct-install lighting projects under MAD 18 for 2021 (Lighting was sampled as one domain and direct install projects were not in the final sample). However, we noted 5 issues with 2021 lighting projects (not direct-install) in which the itemized invoice was not provided and we were not able to evaluate the project due to lack of customer response. In other instances, although we did not have an itemized invoice the customer was able to provide enough information to proceed with evaluation.

- **Mid-Stream Lighting (MAD 44) 2020 Report Recommendation:** "We found that a significant number of products that were incentivized were not on EnergyStar or DLC QPLs, making them ineligible. We did not see documentation of exceptions granted for these products.

  Recommendation: Ensure that all products are on an approved QPL per MAD requirements and document any program exceptions to these requirements in the project's documentation package."

Findings: We did not find any notable issues with DLC QPLs for the 2021 evaluation.

SEM Track

Finding (from the 2019 report) – The Strategic Energy Management program has become a more complicated program over time, which has increased the cost to evaluate the program. The increase in complication is primarily driven by the increase in performance tracking tools (PTTs) used to estimate program savings. While it appears that improvement and consolidation of PTTs is occurring, there are still incidents where model inputs and information are located in inconsistent areas or are not appropriately accounted for in the model.

Recommendation (from the 2019 report) – DNV recommends that Energy Trust continue its efforts to create simplified and consistent PTT tools for program participants to use. DNV recommends the creation of a "Non-Routine Events" (NRE) log within the PTT that documents all capital projects (both those in the baseline and those during program years), any weather adjustments made, and any other NREs that are accounted for in the model (including baseline adjustments and gas curtailments). The log should state how the NRE is accounted for in the savings calculation."

Answer (from SBW 2020 report): A non-routine events log has been added to the Aggregation sheet but it only tags outliers in the baseline where the t-score is greater than 3. We suggest that if using monthly models, the t-score threshold be lowered following Chauvenet's Criterion. We also suggest that the log be used to tag outliers in the reporting year also. Include a test to ensure that values of independent variables in the reporting period do not fall too far beyond the range of values upon which the baseline model was built.

Findings: Our observation for 2020 applies to 2021 as well, except that what we called the "Non-Routine Events Log" in the Aggregation sheet is in actuality the "Baseline Outlier Review" in the current PTT version. The PTT does not include a "Non-Routine Events Log" as described in DNV's 2019 recommendation. We agree with DNV that this should be included in the PTT per our 2020 recommendation.