2014–2017 Industrial and Process Efficiency Program Impact Evaluation

Final Report

Prepared for:

New York State Energy Research and Development Authority

Albany, NY

Dana Nilsson, PhD, PE
Project Manager

Prepared by:

ERS

Corporate Headquarters: North Andover, MA 01845

Nick Collins

Associate Director

Steven Keates

ADM Associates, Inc.

Sacramento, CA 95827-2501

ERS (Principal consulting firm): NYSERDA Contract #104543 September 2018

ADM Associates (Supplemental consulting firm): NYSERDA Contract #104531

NYSERDA Record of Revision

Document Title
2014–2017 Industrial and Process Efficiency Program Impact Evaluation September 2018

Revision Date	Description of Changes	Revision on Page(s)
Add Date	Original Issue	Original Issue

Notice

Table of Contents

List of Figures

List of Tables

1 Introduction

This report describes the impact evaluation of NYSERDA’s Industrial and Process Efficiency Program (IPE program or the Program). The IPE program is a mature industrial resource acquisition program that has been in place since 2009, helping fund a large quantity of custom projects covering many unique and varied measure types. The primary objective of this impact evaluation was to independently estimate the Program’s electric and natural gas energy savings. Electric peak demand impact is not currently required to be reported to the Department of Public Service (DPS), but it is being internally tracked by NYSERDA program and evaluation staff for future consideration. The evaluated savings are based on project-specific measurement and verification (M&V) performed on a statistically representative sample of 55 electric energy savings and 30 natural gas savings projects. Ten projects had a combination of both electric and natural gas savings. Of the sampled projects, 20 projects were deemed a conflict of interest for the principal consulting firm of the Impact Evaluation Team (ERS) due to their involvement in those projects as either the Outreach Contractor, the Technical Reviewer, or the Project Manager. These 20 conflict of interest sites were evaluated by the supplemental consulting firm, ADM Associates. The sampled projects were drawn from a combined total population of 243 electric and natural gas savings projects completed between January 1, 2014, and December 29, 2017, as indicated in Table 1-1, further below.

1.1 Program Description

NYSERDA’s IPE program aims to help manufacturers, data centers, and other production facilities increase output and improve processing as efficiently as possible. NYSERDA calculates financial support for these projects on first-year annual energy savings, and, when appropriate, accounts for reductions in energy intensity (the energy used per unit of production or workload). The projects included the evaluation were initiated through NYSERDA’s Energy Efficiency Portfolio Standard (EEPS-2)[1] and Clean Energy Fund (CEF)[2] Resource Acquisition Transition chapter funding. Table 1-1 outlines the Program’s population of projects for the period of study.

[1] Department of Public Service, Filing #4779 Case No. 07-M-0458, 2016

[2] CEF Quarterly Report, Case 14M0094, 2016

Table 1-1. IPE Program-Reported Savings (1/1/14–12/29/17) for EEPS-2 and CEF Transition Projects

Measure Type/Facility Type	No. of Projects with Completed Measures[a]	Annual Savings	Percentage of Installed Savings
Electric – MWh/yr			
Non-process/all facility types	135	128,811	67%
Process/industrial	39	16,435	14%
Process/data center	29	22,537	20%
Total Electric Savings (MWh/yr)	**188**	**167,783**	**100%**
Natural Gas – MMBtu/yr			
Non-process/all facility types	28	232,689	33%
Process/industrial	27	468,752	67%
Process/data center	0	0	0%
Total Natural Gas Savings (MMBtu/yr)	**55**	**701,441**	**100%**

[a] There were 188 distinct electric projects, but 15 of those projects had measures in more than one Facility Type category.

1.2 Summary of Evaluation Objectives and Methods

The evaluation objectives and methods are summarized in Table 1-2.

Table 1-2. IPE Program Gross Impact Evaluation Objectives and Methods

Objective	Purpose	Method
Estimate gross energy impacts	To establish annualized first-year evaluated gross energy savings based on electric (kWh) and fuel savings (MMBtu) at the customer site	On-site M&V using on-site logging, custom engineering assessments, and/or billing analysis of a representative sample of Program participants.

The objective of this impact evaluation was to estimate the gross savings for the Program, which includes the energy savings for electricity and energy savings for natural gas. The evaluated savings are based on the rigorous project-specific M&V and calculations of representative realization rates (RRs) from a sample of projects from the population. The sample is designed to provide Program gross energy savings with 10% relative precision at 90% confidence for each of three project types (process, data center, and non-process). This report complies with the M&V savings-related requirements listed in *New York Evaluation Plan Guidance for EEPS Program Administrators*,[3] which was issued by the DPS and is intended to provide robust, timely, and transparent results. The methods also comply with the guidelines of the *National Action Plan for Energy Efficiency (NAPEE) Model Energy Efficiency Program Impact*

[3] http://www3.dps.ny.gov/W/PSCWeb.nsf/96f0fec0b45a3c6485257688006a701a/766a83dce56eca35852576da006d79a7/$FILE/NY_Eval_Guidance_Aug_2013.pdf

Evaluation Guide.[4] Overall rigor was high, with site-specific M&V typically following the International Performance Measurement & Verification Protocol (IPMVP) Option A. The Impact Evaluation Team used equipment-specific interval data from evaluation data loggers, customer process management systems, and program-collected data. A few projects were evaluable with production-normalized whole facility utility bill data (Option C). The Impact Evaluation Team later provides the percentage of projects that were evaluated under a given IPMVP option in Table 3-3, found in Section 3.2.2.

The Program has undergone two previous impact evaluations. The results of the previous evaluations are presented in Table 1-3.

Table 1-3. Previous IPE Impact Evaluation Results

Realization Rates	First Evaluation Projects Completed Jan 1, 2009–June 30, 2010	Second Evaluation Projects Completed July 1, 2010–June 30, 2012
kWh RR	0.89	0.91
MMBtu RR	1.14	0.96

[4] http://www.epa.gov/cleanenergy/documents/suca/evaluation_guide.pdf

2 Results, Findings, and Recommendations

This section presents the results and findings of the gross impact evaluation analysis and provides recommendations for Program considerations based on those findings.

2.1 Results

The results and findings of the impact evaluation are presented below, broken out by fuel type (electric and natural gas). Through the gross impact evaluation, the Impact Evaluation Team found strong realization rates for both types and found that projects that received Program M&V had higher realization rates than those projects that did not receive Program M&V. The differences analysis (Section 2.2.2) found that there were no systemic issues driving the differences between the reported and evaluated savings, and that a large portion of the differences (approximately 30% of the kWh difference and 53% of the MMBtu difference) are due to changes in facility or system operation that took place after Program involvement, represented the new normal operating conditions, and were not possible for the Program to anticipate.

2.1.1 Electric Results

The realization rate (RR) for the electric energy savings component of the Program is 0.86. Table 2-1 provides the results of the evaluation effort for electric savings.

Table 2-1. Summary of Reported and Evaluated Savings for EEPS-2 and CEF Transition Funded Projects

Electric Savings, MWh	
Total reported savings	167,783
Evaluated gross savings	144,916
Weighted RR	86%
Error ratio	0.34
Standard deviation	0.38
Project population	188

Table 2-2 details the evaluated electric savings, broken out by facility type, and includes additional information including sample size, relative precision, and error ratio.

Table 2-2. Electric Savings (MWh) Overall and By Facility Type for EEPS-2 and CEF Transition Funded Projects

Metric	All Electric Projects	Industrial Processes	Data Center Processes	Industrial and Data Center Non-Processes
Sample size	55	17	16	22
Population size	188	39	29	135
Weighted tracking savings, MWh	171,336.5	23,712.6	26,673.4	120,950.4
Weighted evaluated savings, MWh	147,984.8	20,353.7	29,069.7	98,561.4
Total reported savings, MWh	167,782.9	16,434.7	22,537.4	128,810.8
Evaluated gross savings, MWh	144,915.6	14,106.7	24,562.1	104,966.8
Weighted RR	86%	86%	109%	81%
Standard error	3.9%	3.5%	3.4%	5.4%
Relative precision at 90% confidence	7.4%	6.7%	5.2%	11.0%
Standard deviation of the RR	0.38	0.38	0.36	0.35
Error ratio	0.34	0.31	0.28	0.35
% of evaluated gross savings	100%	14%	20%	67%

While the sample was designed to achieve 90/10 confidence and relative precision, the non-process group achieved a relative precision of 11%. This is driven by the greater-than-anticipated variability in RR and the lower weighted RR for this group. The anticipated variability (error ratio) of this stratum as used in the sample design was 0.30. The resulting error ratio based on evaluation findings is 0.35. This greater-than-anticipated variability is reflected in the standard error of this stratum, which, when multiplied by the z-value of 1.645 for 90% confidence, results in the numerator of the relative precision calculation. The denominator of this calculation is the weighted RR, resulting in the "relative" precision. Further detail on sample design and weighting is included in Appendix A.

Figure 2-1 presents a scatter plot of the reported savings and the evaluated savings. The diagonal line in the graph represents an RR of 1.0. If projects achieve an RR of 1.0, they will fall directly on this line. Projects with an RR greater than 1.0 are above this line; projects with an RR less than 1.0 are below this line. The axis maximum values have been limited to preserve visual clarity, which prevents one large project (reported savings= 23,224,988 kWh; evaluated savings= 19,919,689 kWh) from appearing on this graph. Additional information on the source of differences between the reported and evaluated savings values is presented in Section 2.2.2.

Figure 2-1. Plot of Reported vs. Weighted Evaluated Electric Savings for EEPS-2 and CEF Transition Funded Projects

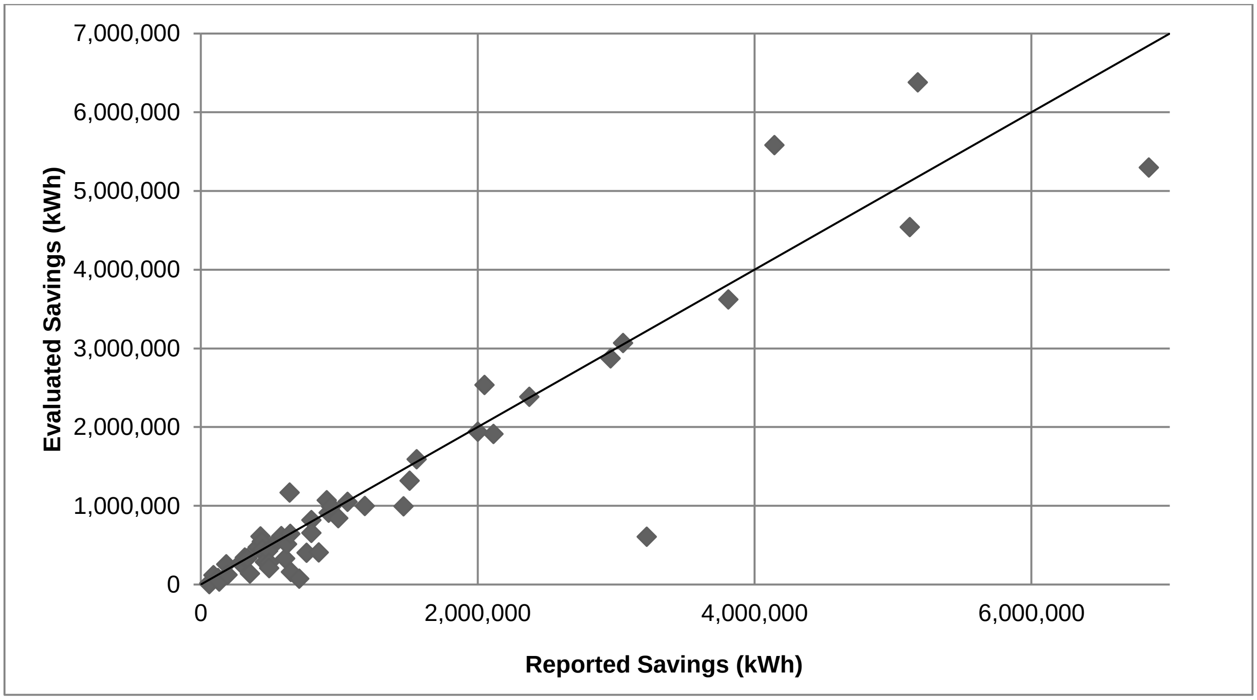
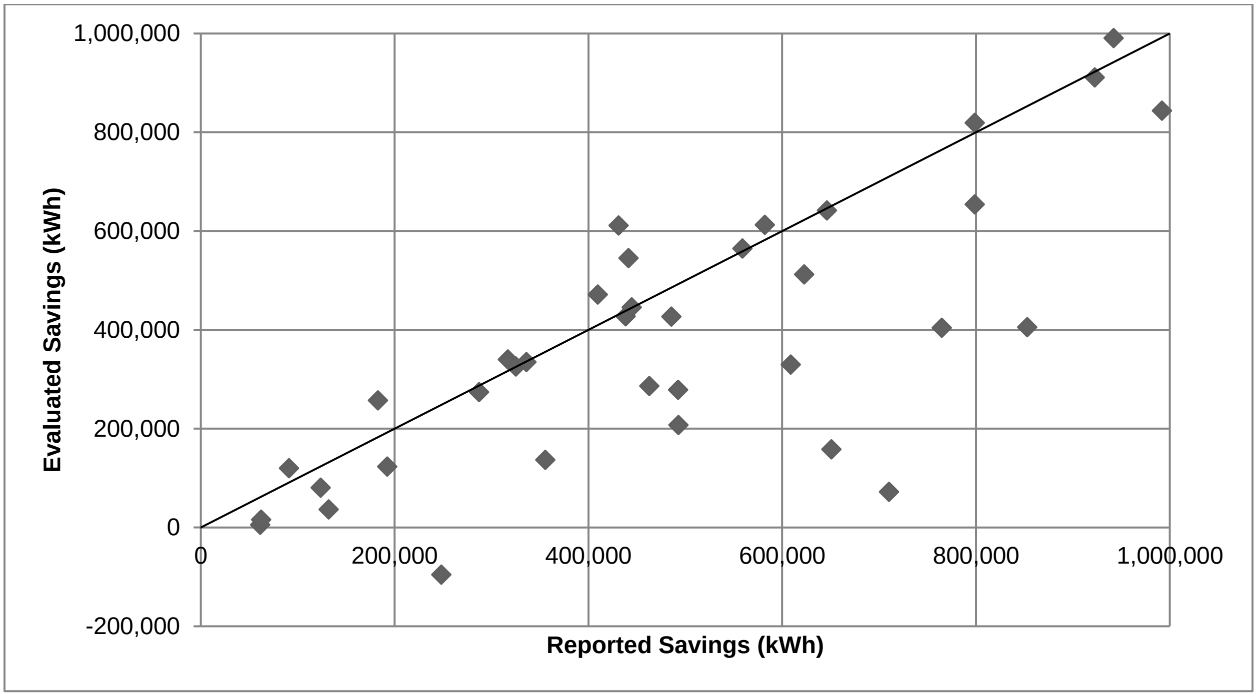


Figure 2-2, below, presents the same information, but with the axes limited further to provide more granularity on the performance of smaller projects.

Figure 2-2. Focused Plot of Reported vs. Weighted Evaluated Electric Savings for EEPS-2 and CEF Transition Funded Projects Under 1,000,000 kWh



2.1.2 Natural Gas Results

The RR for the natural gas savings component of the Program is 0.91. Table 2-3 provides the results of the evaluation effort for natural gas savings.

Table 2-3. Summary of Reported and Evaluated Savings for EEPS-2 and CEF Transition Funded Projects

Natural Gas Savings	
Total reported savings, MMBtu	701,441
Evaluated gross savings, MMBtu	635,675
Weighted RR	91%
Relative precision	1.5%
Error ratio	0.42
Projects population	55

Table 2-4 details the weighted evaluated natural gas savings broken out by facility type, and includes additional information including sample size, relative precision, and error ratio.

Table 2-4. Natural Gas Savings Overall and By Facility Type for EEPS-2 and CEF Transition Funded Projects

Metric	All Natural Gas Projects	Industrial Processes	Industrial and Data Center Non-Processes
Sample size	30	16	14
Population size	55	27	28
Weighted tracking savings, MMBtu	701,646	475,182	226,464
Weighted evaluated savings, MMBtu	635,861	439,768	196,093
Total reported savings, MMBtu	701,441	468,752	232,689
Evaluated gross savings	635,675	433,817	201,483
Weighted RR	91%	93%	87%
Standard error	0.8%	1.1%	1.1%
Relative precision	1.5%	2.0%	2.1%
Standard deviation of the RR	0.35	0.33	0.39
Error ratio	0.42	0.47	0.38
% of evaluated gross savings	100%	69%	31%

Figure 2-3 presents a scatter plot of the reported savings and the evaluated savings. The diagonal line in the graph represents an RR of 1.0. If projects achieve an RR of 1.0, they will fall directly on this line. Projects with an RR greater than 1.0 are above this line; projects with an RR less than 1.0 are below this line. Additional information on the source of differences between reported and evaluated savings values is represented in Section 2.2.2.

Figure 2-3. Plot of Reported vs. Weighted Evaluated Natural Gas Savings for EEPS-2 and CEF Transition Funded Projects

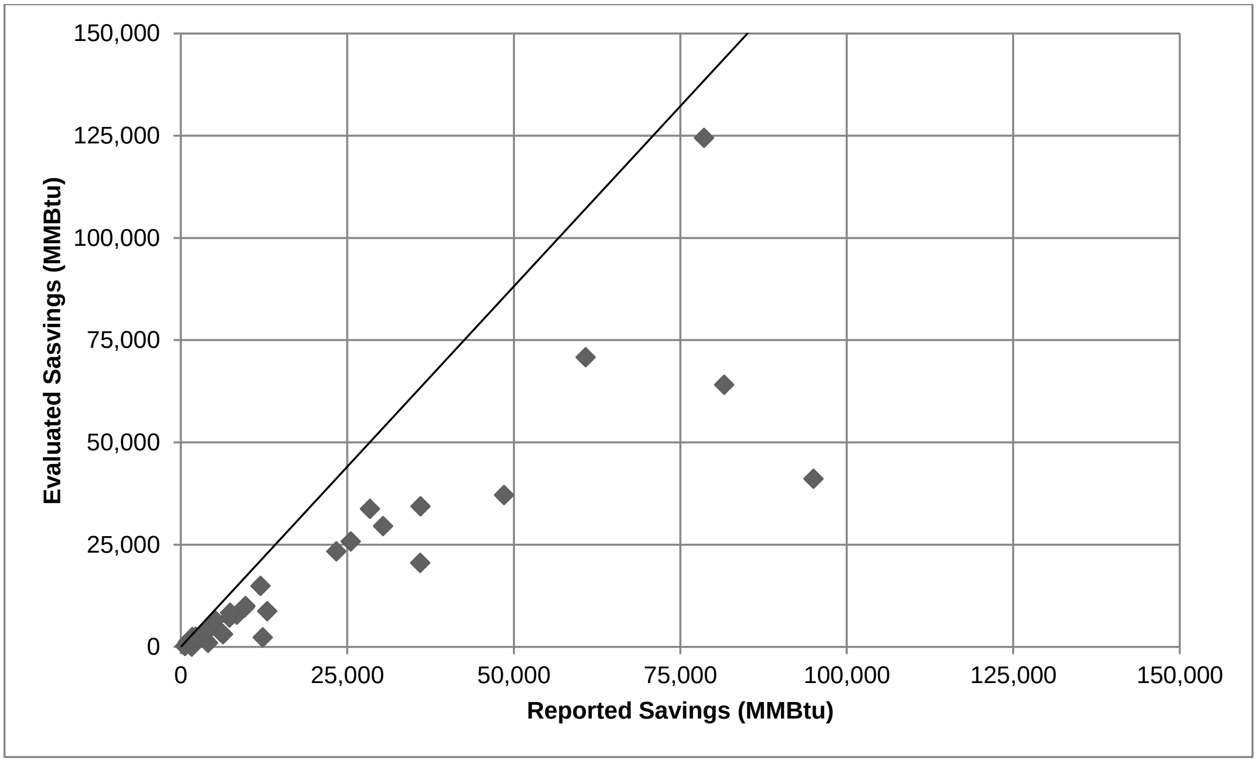
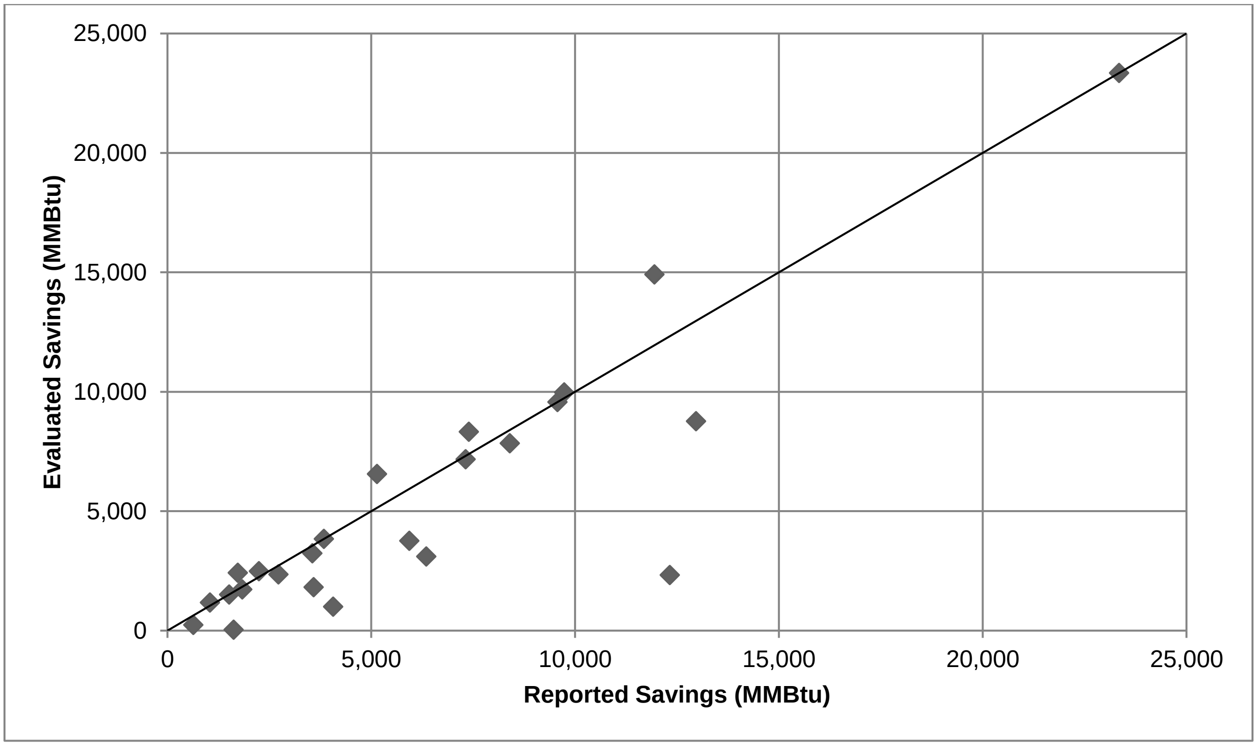


Figure 2-4 presents the same information, but the axes have been limited to provide more granularity on the performance of smaller projects.

Figure 2-4. Focused Plot of Reported vs. Weighted Evaluated Savings for EEPS-2 and CEF Transition Funded Projects Under 25,000 MMBtu



2.2 Findings

The analysts dissected the results along several boundaries to discern patterns and better understand why RRs varied from 1.0. Stratification criteria included measure type, measure size, and whether Program M&V had been performed. Additional information on sample design is provided in Section 3 and Appendix A. The discussion on the differences analysis findings (Section 2.2.2) goes beyond the numerical results and presents a quantitative and qualitative assessment of the factors that resulted in differences between the reported and evaluated savings.

The findings in this section are presented without regard to sample weights. Since the categories used in the following breakouts were not used as strata within the sample design, the results in this section only represent the sample (not the population at large). While these findings may not be statistically applicable to the overall Program, they are nonetheless informative relative to this statistically representative sample. Further, the unweighted values presented in this section will not align with or sum to the final weighted values presented in Section 2.1.

2.2.1 Aggregate Analysis Findings

This section details the results of the evaluation at the measure level and includes an examination of the different characteristics for more insight on the findings.

Measure-Level Results

This section presents the performance of the sample at the measure level. Many projects include more than one measure type, but for tracking purposes NYSERDA categorizes each project with only one measure type. The Impact Evaluation Team broke out each measure type and its related savings at the project level to provide a more detailed view into measure-level performance within the sample. Figures 2-5 and 2-6 present the quantity of each measure type within the electric sample and the average unweighted RR of the measure type, respectively.

Figure 2-5. Quantity of Measure Types within the Electric Sample

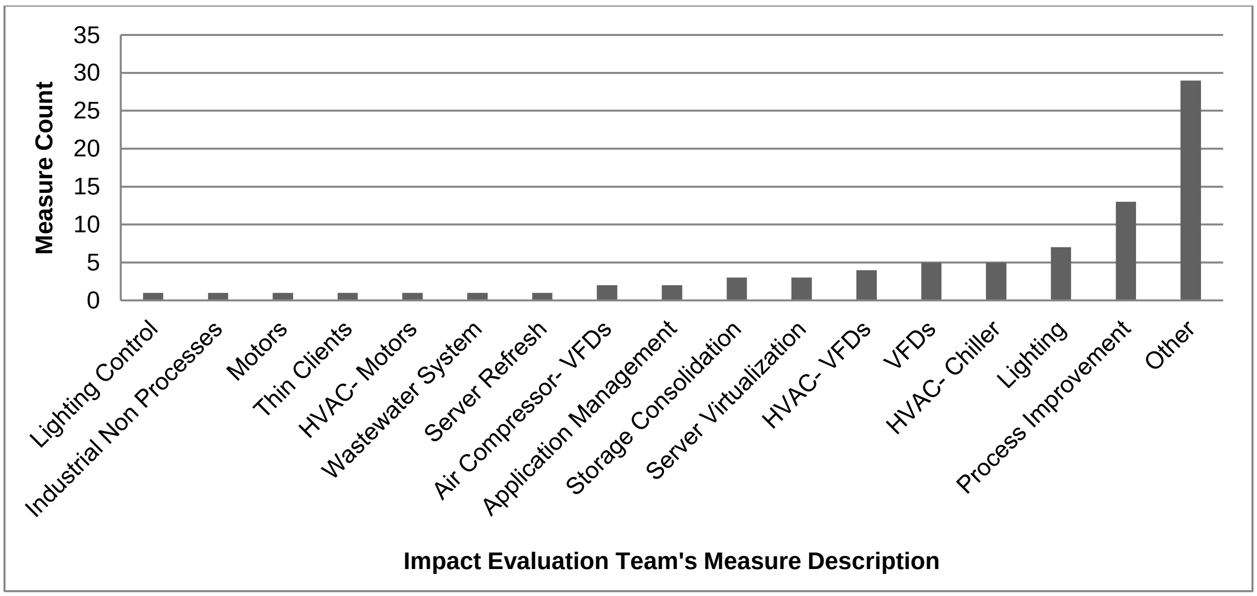
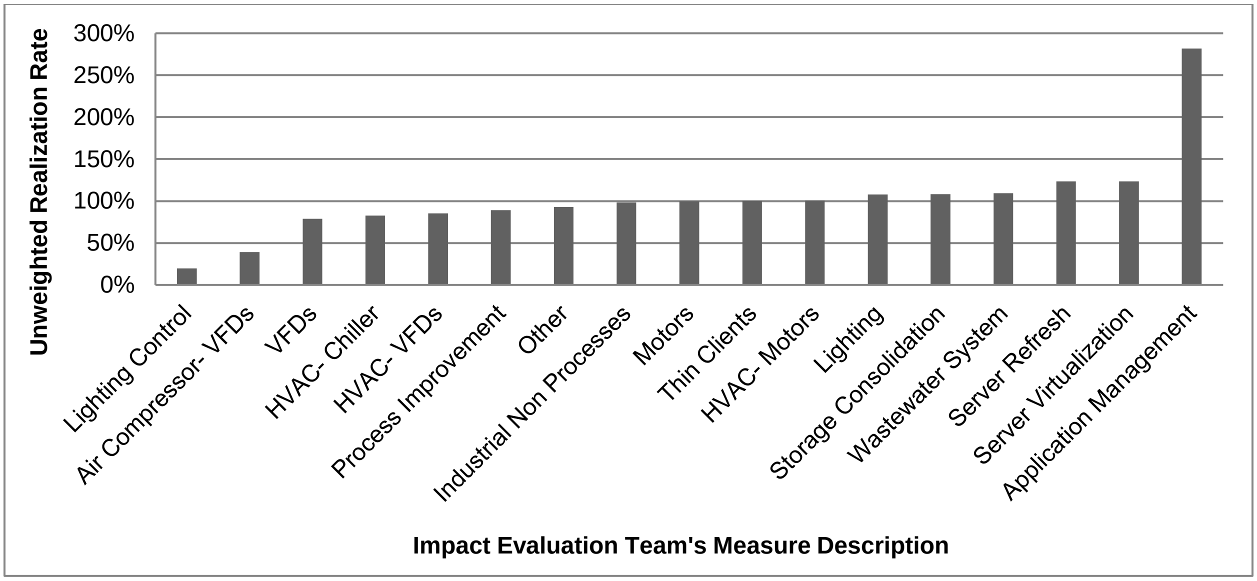


Figure 2-6. Average Unweighted RR by Measure Type for the Electric Sample



Figures 2-7 and 2-8 present measure quantities and average unweighted RRs for the natural gas sample.

Figure 2-7. Quantity of Measure Types within the Natural Gas Sample

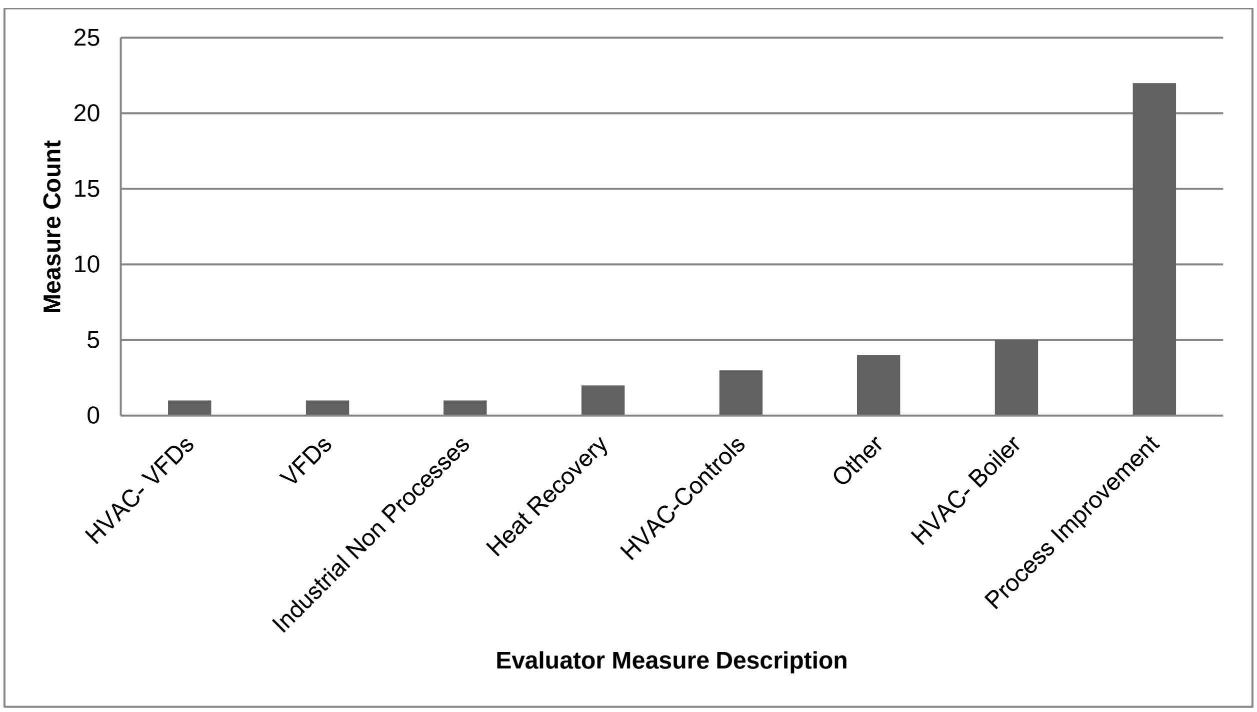
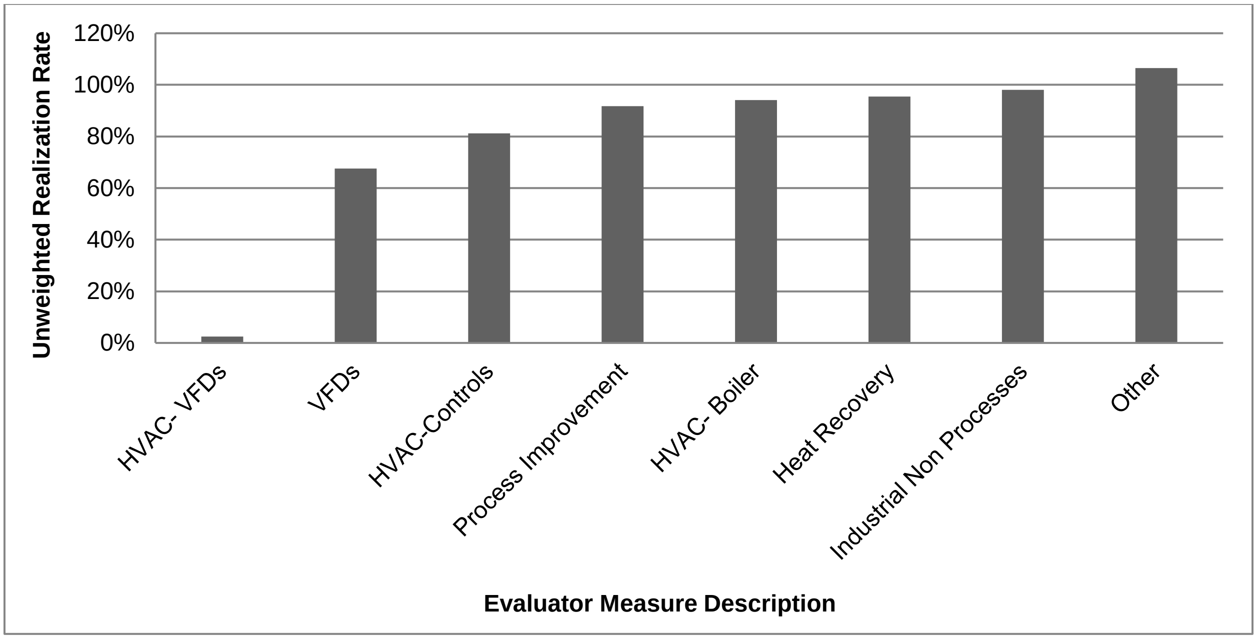


Figure 2-8. Average Unweighted RR by Measure Type for the Natural Gas Sample



2.2.1.1 Program M&V Influence

The Impact Evaluation Team analyzed the evaluation findings in terms of whether or not Program M&V had been performed. The IPE program requires that M&V be performed by the applicant for projects whose savings exceed a certain threshold.[5] Table 2-5 presents the RRs for electric and natural gas projects with and without Program M&V.

Table 2-5. Electric and Natural Gas Unweighted Realization Rates Dependent on Program M&V

Fuel Type	With Program M&V	Without Program M&V
Electric, kWh	96%	81%
Natural gas, MMBtu	93%	83%

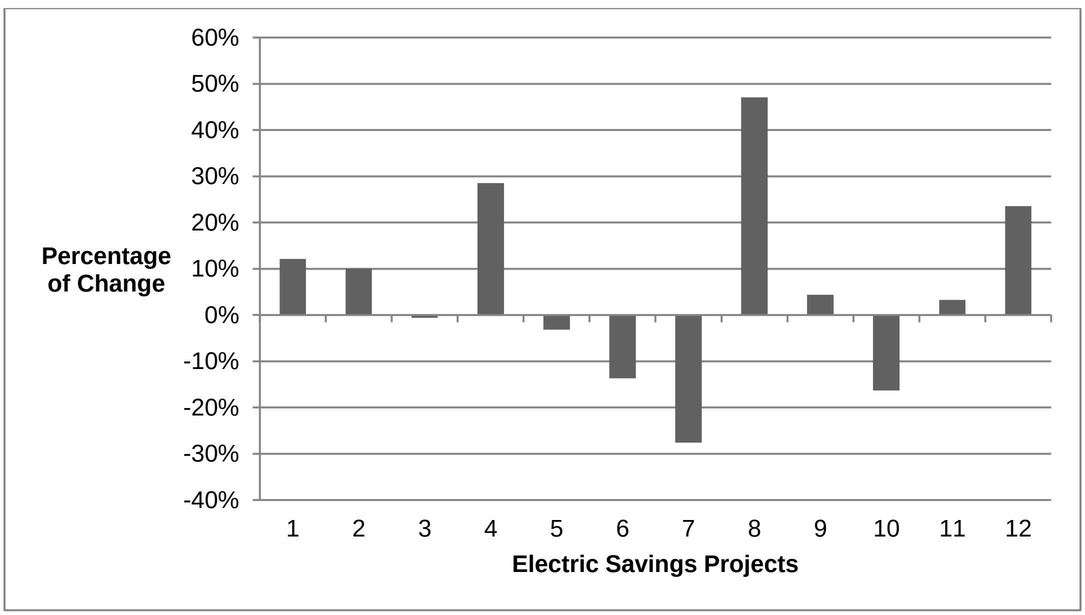
Savings are calculated up to three times during NYSERDA's project review process: once during the Engineering Analysis (EA) phase and again at project completion during the Project-Installation Review (PIR); and then for projects requiring Program M&V, a third time at the conclusion of an M&V period. Projects that receive Program M&V are larger and typically more complex, thus lending themselves to greater uncertainty in the program-reported project calculations. NYSREDA's third-party Technical Reviewers plan and conduct the M&V activities in concert with NYSERDA program staff. Program

[5] See Appendix C for Program M&V thresholds.

M&V for these larger, more complex projects helps establish accurate baselines, efficient-case equipment performance, and production rates. Select projects that are particularly large or technically complex may undergo a concurrent evaluation process where NYSERDA Program staff, Technical Reviewers, and an additional third-party independent evaluator collaborate to characterize the baseline systems and performance and to generate M&V plans.

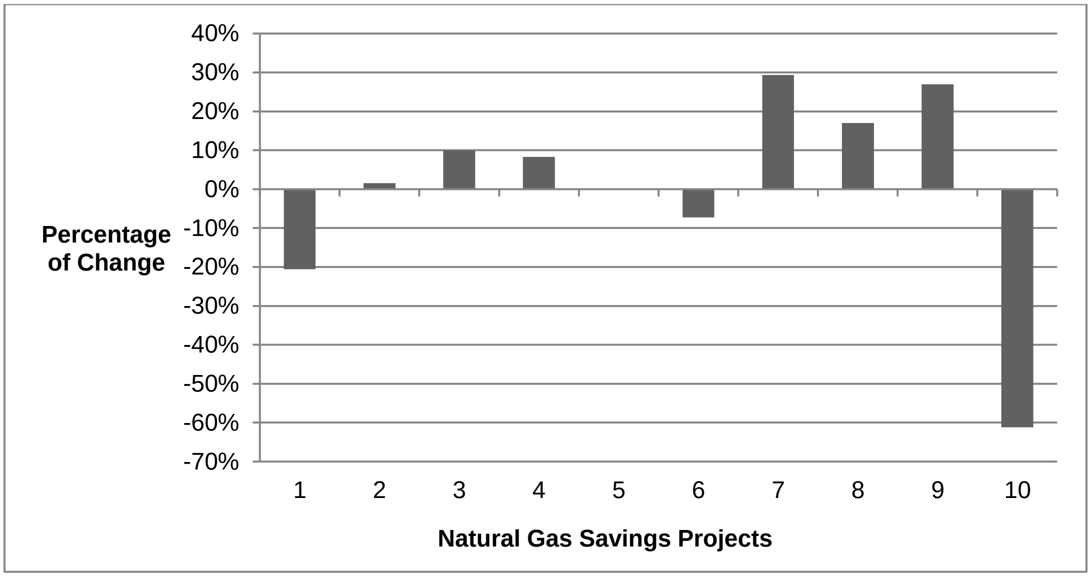
There were six projects within the sample that had received concurrent evaluation. Projects receiving Program M&V almost always have their savings revised as a result of the M&V effort, which has led to the strong unweighted RRs reported in Table 2-5, above. The strong RRs of projects receiving Program M&V is indicative of successful and critically important Program M&V; without Program M&V, reported savings would be less precise. A review of the difference between PIR and M&V phase savings yields an average absolute change between the two phases of 17%. Figures 2-9 and 2-10, below, detail the percentage of change in Program-reported savings between PIR and M&V phases for electric and gas, respectively. Within the sample, there were 12 electric projects with distinct reported savings between PIR and M&V reports, and 10 natural gas projects with reported savings at both the PIR and M&V reports.[6] The remaining sampled projects that received Program M&V submitted a combined final PIR/M&V report that reported only the final M&V based savings.

Figure 2-9. Percentage of Change in Program-Reported Savings between PIR and M&V Phase for kWh Projects



[6] There are a total of 23 electric projects and 12 natural gas projects in the sample that received Program M&V.

Figure 2-10. Percentage of Change in Program-Reported Savings between PIR and M&V Phase for NG Projects



Achieving RRs of 90% or greater may be expected of a Program with such rigorous M&V requirements, but it is a testament to the competency of Program staff and Technical Reviewers that RRs of 81% and 83% can be achieved through ex-ante[7] calculations alone. The results of this evaluation are similar to the previous IPE impact evaluation in terms of the RRs of projects that did and did not receive Program M&V. Table 2-6 present the RRs based on Program M&V for the current and previous impact evaluations.

Table 2-6. Current and Previous Evaluation RRs by Program M&V

Realization Rates	With Program M&V		Without Program M&V	
	Previous Evaluation	Current Evaluation	Previous Evaluation	Current Evaluation
kWh RR	92%	96%	85%	81%
MMBtu RR	97%	93%	93%	83%

By comparison, recent gross impact evaluations of large custom industrial programs in Massachusetts and California yielded lower RRs then those found in the current or previous IPE evaluations. Table 2-7 provides the RRs from those recently published studies.

[7] Based on forecasts rather than measured results

Table 2-7. Recent National Industrial Program Impact Evaluation Results

Realization Rates	Previous IPE Evaluation	Current IPE Evaluation	Massachusetts C&I Impact Evaluation of 2013 Custom Process[a]	California 2013–2015 Custom Impact Evaluation[b]	California 2010–2012 Annual Progress Evaluation Report Industrial Findings[c]
kWh RR	91%	86%	63%	44%–66%	60%–70%
MMBtu RR	96%	91%	N/A	50%–63%	60%–70%

[a] Massachusetts Commercial and Industrial Impact Evaluation of 2013 Custom Process Installations, published 2017: http://ma-eeac.org/wordpress/wp-content/uploads/MA-2013-CI-Custom-Process-Impact-Evaluation.pdf

[b] California 2013–2015 Custom Impact Evaluation Results, published 2017: http://www.calmac.org

[c] California 2010–2012 Energy Efficiency Annual Progress Evaluation Report, published 2015, industrial findings: http://www.cpuc.ca.gov/General.aspx?id=6391

2.2.2 Differences Analysis Findings

This section summarizes the differences in savings in the sampled projects and discusses the potential areas for improvement. The findings are applicable to both EEPS-2 and CEF projects, as projects of both funding sources were evaluated and are very similar. Some of the differences observed are from EEPS-2-funded projects whose measure type is no longer offered. The intent of documenting these differences is to report the best possible estimate of grid impact. While some differences are not directly attributable to the action or lack of action of the Program – such as when a load shape has changed after Program involvement – they are still accounted for to accurately represent the impacts of these industrial efficiency projects.

A review of the differences between the program-reported and evaluated savings demonstrates some findings that are valuable for improvement of the program. Figures 2-11 and 2-12 break down the sum of positive and negative differences found in sampled projects for each fuel. They are separated into three primary categories, corresponding to the program phases of Engineering Analysis, Project Installation Review, and Measurement & Verification, representing the Impact Evaluation Team's understanding of when the difference originated. For example, a difference listed under Application Review originated before the project was installed and could potentially have been corrected by the applicant or program staff at the time of the EA Report, whereas those differences listed in Measure Performance occurred after installation and post-inspection.

Figure 2-11. Unweighted Electric Savings Differences by Project Phase

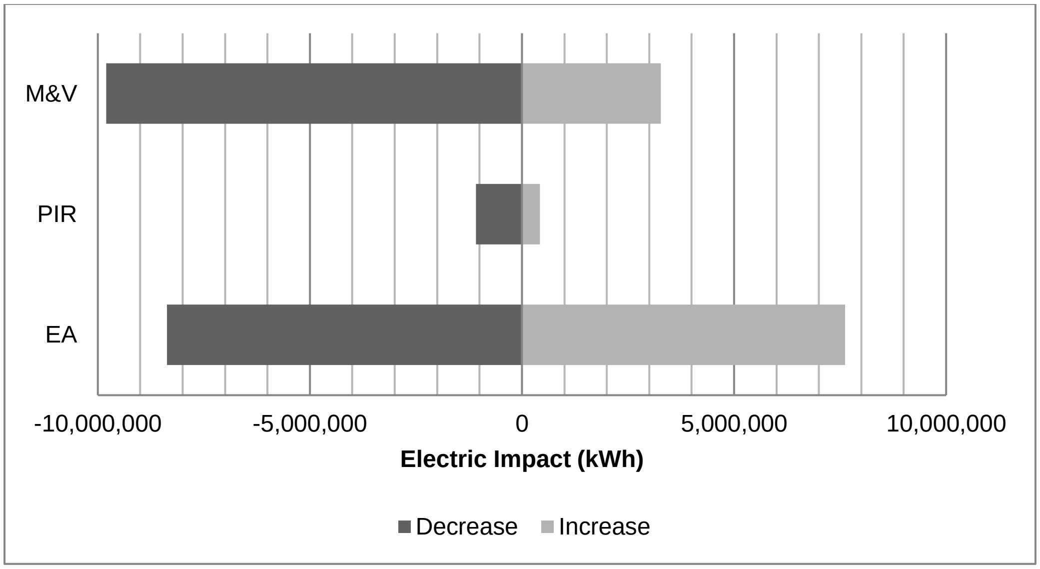
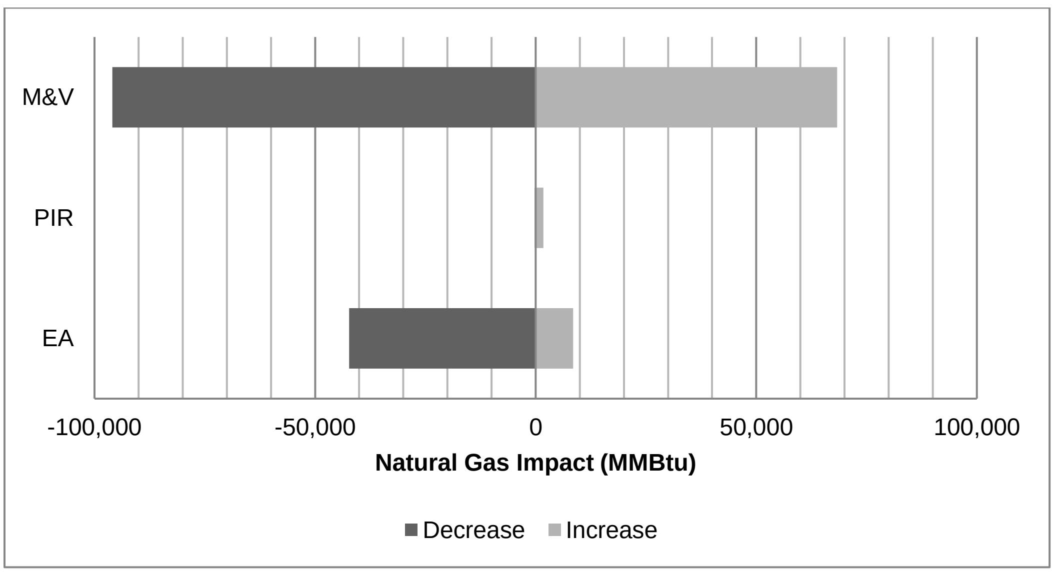


Figure 2-12. Unweighted Natural Gas Savings Differences by Project Phase



These figures reveal that the majority of savings differences occurred at Measure Performance, after installation of the project. It is not surprising for a mature and well-functioning program that the majority of differences were attributable to as-built performance. For projects that received Program M&V, many of these differences are driven by changes that took place after Program M&V (e.g., a different load profile that represents a new typical operation) and could not have been predicted by the Program. The Application Review category also shows significant differences, and these should be the target for improvement in the future, as they are attributable to issues in the savings calculations that were likely preventable.

Tables 2-8 and 2-9, below, show the differences in the same three primary categories and in subcategories of similar issues, for electric and gas measures, respectively. An actual in-project example is listed to illustrate the type of difference found in that subcategory, alongside the total number and total unweighted impact of positive and negative differences. A positive difference refers to an instance in which the Impact Evaluation Team made a change to the savings calculation that led to higher savings on the measure than the program-reported value. A negative difference refers to an instance in which the Impact Evaluation Team made a change to the savings calculation that led to lower savings on the measure than the program-reported value.

The differences are dispersed among numerous categories, and while certain categories represent a larger portion of the differences (changes in load profile for example), it is important to note that no systemic differences were found. Differences, even within a given category, were diverse in nature and reflect the unique nature of the projects and the TRs approach to estimating energy savings.

Table 2-8. Key Drivers in Electric Savings Differences

Key Drivers in Electric Savings Differences with Site-Specific Examples			Negative Differences		Positive Differences	
Difference Primary Category (Project Phase)	Difference Sub-Category	Actual Project Example	# Projects	Impact on kWh	Impact on kWh	# Projects
Application review (EA)	Difference in as-built equipment efficiency	The applicant used a hardcoded EER value of unknown origin. Evaluators revised the AHRI EER and NPLV based on the glycol solution in the system.	7	-194,319	12,719	3
	Inaccurate estimation from applicant model	The applicant model did not account for increased usage of the chilled water pump.	18	-952,343	3,390,965	11
	Inaccurate normalization to typical weather	The applicant used historical weather data of a weather station which is the closest weather station to the facility. To represent the project annual savings of a typical year, the evaluators used TMY3 weather data of the same station.	3	-113,553	54,099	4
	Incorrect baseline reference	The applicant analysis uses the existing chiller efficiency as the baseline chiller efficiency, while the ex post uses a code minimum efficiency baseline since the project was a capacity expansion.	10	-6,209,059	2,493,670	10
	Insufficient assessment of measure interactivity	The applicant did not include cooling interactivity when calculating savings from reduced data center server power.	5	-829,201	1,478,272	6
	Difference in equipment hours of operation	Applicant analysis did not account for presence of the existing free cooling system, which led to reduced annual chiller run hours in the evaluator savings calculation.	3	-69,288	187,968	1
Measure installation verification (PIR)	Difference in cooling or heating interactivity	The applicant did not include UPS efficiency in chiller interactivity calculation, resulting in greater cooling interactivity in the evaluator calculation.	1	-11,763	67,330	1
	Difference in quantity installed	The project removed only two CRAH units instead of all three denoted by the applicant.	2	-14,970	175,341	1
	Difference in installed control strategy	The applicant's savings model did not match the installed control strategy employed onsite. The evaluators found that the as-built fan VFD was set to a fixed speed of 100%.	4	-535,899	179,647	1
	Difference in installed equipment size	The applicant modeled savings based on a smaller heating recirculation pump motor than was reported installed by the site staff.	1	-35,963	0	0
	Inoperable measure	The measure was not installed.	4	-489,918	0	0
Measure performance (M&V)	Difference in analysis methodology	The applicant's analysis method applied percent full-load amps (FLA) to motor curve data and thus, did not account for power factor. This approach results in greater savings than the evaluator analysis because the % FLA is reduced post-retrofit, but that does not equate to reduced power demand.	1	-548,847	196,274	4
	Difference in as-built equipment efficiency	The applicant assumed 98% UPS efficiency using an assumed loading, and the observed load changed, resulting in 96.4% efficiency according to manufacturer's documentation.	3	-169,330	204,136	7
	Difference in cooling or heating interactivity	The applicant used the chiller plant full load kW/ton for the interactive cooling savings, where the plant load is best represented by an IPLV efficiency.	2	-174,538	326,842	2
	Difference in equipment hours of operation	The evaluators used the metered data to calculate the operating hours of the as-built compressed air dryer. The evaluated operating hours were lower than the applicant's value, resulted in lower savings than the reported savings.	13	-986,335	1,207,550	8
	Difference in equipment load profile	The evaluators used metered data and spot measurements to determine that the average motor load was lower than expected in the applicant analysis.	19	-7,613,832	1,336,206	15
	Inaccurate estimation from applicant model	The applicant used an average operating kW with an assumed annual hours to determine savings. The evaluators used average weekly profiles to determine annual kWh, which account for variations in loading and efficiency more precisely, and found a lower demand than the applicant.	5	-308,004	0	0
Totals			101	-19,257,162	11,311,020	74

Table 2-9. Key Drivers in Gas Savings Differences

Key Drivers in Gas Savings Differences with Site-Specific Examples			Negative Differences		Positive Differences	
Difference Primary Category (Project Phase)	**Difference Sub-Category**	**Actual Project Example**	**# Projects**	**Impact on MMBtu**	**Impact on MMBtu**	**# Projects**
Application review (EA)	Inaccurate estimation from applicant model	The applicant had a calculation mistake that added the tonnage saved for water to the plant efficiency and then multiplied that by the tonnage saved with the upgrade. The correct equation is to add the two improvements tonnage saved together and then multiply that by the plant efficinecy.	13	-34,280	4,808	5
	Inaccurate normalization to typical weather	The applicant based the run hours on a weather station that was not the closest and most representative weather station.	0	0	1,646	1
	Incorrect baseline reference	The applicant accounted for product loss avoidance in the savings analysis because the new equipment would function more reliably and not spoil any product. While this is true, the evaluated baseline does not allow this to account for savings, and the baseline should show the same production level as the efficient case.	1	-12	1,417	1
	Unknown applicant algorithm or assumptions	Applicant savings analysis was performed by a third party contractor and the calculation algorithm inputs were only partially provided in pdf format. The evaluators determined savings using a steam trap savings tool, using the known inputs from the applicant were used in the calculation, and manufacturer's orifice sizes where the applicant data gave no information.	1	-2,214	615	1
	Insufficient assessment of measure interactivity	The applicant calculations assumed that the plant efficiency would remain the same despite other measures in the project affecting plant efficiency.	1	-5,337	0	0
	Ineligible measure	The measure is duplicate savings claim from other measures	1	-428	0	0
Measure installation verification (PIR)	Difference in installed control strategy	The evaluators found that the hot water supply temperature setpoint was lower than thte assumed value, and according to site staff had been that way since installation. This resulted in increased savings.	0	0	17	2
	Difference in as-built equipment efficiency	The applicant boiler efficiency in the savings calculation did not match verified equipment specifications.	1	-157	0	0
	Difference in quantity installed	The applicant post inspection revealed that less steam traps were verified to be installed than the number used in the calculation of savings.	2	-63	1,712	1
Measure performance (M&V)	Difference in analysis methodology	The applicant used an Option A approach, and evaluators used a billing analysis (Option C) to be more certain of measure interactivities.	4	-4,901	1,098	2
	Difference in as-built equipment efficiency	The applicant as-built efficiency, in pounds steam per product, was lower than the efficiency measured by evaluator M&V.	3	-3,411	13,055	3
	Difference in equipment hours of operation	The applicant annual production estimate did not account for downtime periods. Based on the production data provided by the site contact, the as-built equipment press has a downtime period of 51% of the total operating hours. The evaluated annual production was lower as a result.	7	-6,238	343	2
	Difference in equipment load profile	The applicant used a binned temperature analysis and assumed the boiler was fully loaded during the shoulder seasons. The evaluator's billing analysis determined savings based on actual gas in these periods was less than the modeled full load.	4	-60,218	53,827	6
	Difference in cooling or heating interactivity	The applicant used multiple independent measure calculations to estimate savings. The facility shut down permanently, making a normalized billing analysis (IPMVP Option C) the only evaluation method available, and the presumed reason for difference is inadequate accounting of the heating interactivity between the measures of the project.	2	-16,943	0	0
	Difference in installed control strategy	The applicant modeled savings based on the planned control strategy of modulating supply air fan speed based on certain exhaust air flows. When evaluators collected trends of the equipment and dsicussed with teh site contact, it was clear that the planned controls had failed, and a less sophisticated control strategy was implemented to keep the air handler operating without errors.	1	-4,207	0	0
Totals			41	-138,409	78,538	24

A small number of high-impact differences occurred in the electric sample, which are helpful to examine in tandem with Table 2-8 to fully understand what drove the differences. Table 2-10, below, shows the 10 largest electric differences. Note that 7 of these 10 are related to supercomputer measures.

Table 2-10. Largest Single Differences in Electric Savings Measures

Primary Category	Subcategory	Description of Difference	Unweighted Impact kWh
Measure performance	Difference in equipment load profile	The as-built supercomputer was operating at a lower capacity than was demonstrated during Program M&V.	-3,390,890
Application review	Incorrect baseline reference	The assumed baseline for the lighting controls project was that existing lights were on 8,760 hours. Trend data from the customer indicated lower baseline hours.	-2,617,443
Measure performance	Difference in equipment load profile	The as-built supercomputer was operating at a lower capacity than was demonstrated during Program M&V.	-2,300,599
Application review	Incorrect baseline reference	The assumed baseline efficiency (kW/TF) for the supercomputer was adjusted based on new data published on Top500.org for the relevant baseline year.	-1,700,276
Application review	Inaccurate estimation from applicant model	The submitted savings model for the supercomputer did not use a demand curve that included an idle energy demand. When the large data center operates at low loads, the submitted power curve model outputs a lower power consumption than actually occurs.	1,367,806
Application review	Incorrect baseline reference	The assumed baseline efficiency (kW/TF) for the supercomputer was adjusted based on new data published on Top500.org for the relevant baseline year.	1,261,828
Application review	Inaccurate estimation from applicant model	The submitted savings model for the supercomputer did not use a demand curve that included an idle energy demand. When the large data center operates at low loads, the submitted power curve model outputs a lower power consumption than actually occurs.	1,105,210
Application review	Incorrect baseline reference	The assumed baseline efficiency (kW/TF) for the supercomputer was adjusted based on new data published on Top500.org for the relevant baseline year.	-988,045
Application review	Insufficient assessment of measure interactivity	The submitted savings for data center IT load reduction measures did not include cooling interactivity.	986,123
Measure performance	Difference in equipment hours of operation	Production rates were higher than predicted, which led to higher operating hours and higher savings.	665,228

Table 2-11, below, summarizes the same information for gas savings.

Table 2-11. Largest Single Differences in Gas Savings Measures

Primary Category	Subcategory	Description of Difference	Unweighted Impact MMBtu
Measure performance	Difference in equipment load profile	The observed load and as-built efficiency of the process improvement were lower than anticipated.	-53,887
Measure performance	Difference in equipment load profile	Unanticipated increases to production led to increased savings.	45,928
Application review	Inaccurate estimation from applicant model	The submitted savings calculation model did not accurately represent the measure and conditions observed.	-17,527
Measure performance	Difference in as-built equipment efficiency	The process had higher efficiency (steam use per unit production) than anticipated, resulting in higher savings.	11,859
Measure performance	Difference in cooling or heating interactivity	A normalized billing analysis approach revealed that the submitted bottom-up estimates of interactive heating effects in the building were not sufficiently accounted for.	-11,445
Application review	Inaccurate estimation from applicant model	The submitted savings calculation had an algorithm error that added avoided plant tonnage to the plant efficiency and then multiplied that by the tonnage saved. The correct equation is to add the two tonnage-saved values together and then multiply that by the plant efficiency.	-8,384
Measure performance	Difference in cooling or heating interactivity	A normalized billing analysis approach revealed that total savings from the measures was lower than anticipated. The facility was shut down prior to evaluation, and evaluator's best estimate to the cause of the difference is due to interactive effects, since no changes occurred to production, hours, or occupancy between the two analysis approaches.	-5,498
Measure performance	Difference in equipment load profile	Based on facility trended data, it was verified that the facility operates airflow at a lower rate (50%) using the VFD and for a longer period of time (7,000 hours/year) compared to submitted estimates (90% flow rate for 3,744 hours/year), which increases savings.	5,366
Application review	Insufficient assessment of measure interactivity	The submitted savings include several measures that affect plant efficiency, but all individual measures were analyzed with the existing plant efficiency as the baseline, rather than a parametric run approach that included iterative improvement to the plant efficiency from the first measure into the second, and so on. This led to double counting of savings.	-5,337
Measure performance	Difference in equipment load profile	The applicant used a binned temperature analysis and assumed the boiler was fully loaded during the shoulder seasons. The evaluator's savings that were determined by the billing analysis (based on actual gas in these periods) were lower than the modeled full load.	-4,894

When examining the broad trends identified in Tables 2-8 and 2-9, and accounting for the most significant errors shown in Tables 2-10 and 2-11, it becomes clear that while some categories of differences have had the highest effect, they are generally driven by a handful of significant differences rather than a far-reaching and consistent trend of issues. Where there are a high number of differences – such as the 34 in the Measure Performance: Differences in Load Profile category for electric measures – it should be noted that a majority of the savings impact comes from just a few instances, and the majority of the differences were small in magnitude.

Differences that occur during Application Review are the most actionable type of difference, and they are primarily comprised of baseline characterization differences and inaccurate estimations within savings calculations. Baseline items included both baseline characterization issues – like using existing conditions for a project that included capacity expansion – and more specific baseline-related issues – like using assumed efficiency values instead of available equipment specifications. Inaccurate estimation is a broad subcategory that encompasses inaccurate models, unsupported assumptions, and calculation errors. In all cases, the errors persisted through the final Program savings calculation. No patterns were revealed in the analysis of realization rates as a function of the technical reviewer firm.

There were very few errors that occurred at the time of project installation. This indicates that, generally speaking, potential changes in the project at the time of installation are being captured effectively by the post-installation inspection.

As noted above, many differences (both positive and negative) happened fully after any involvement with the Program, even for projects that received Program M&V. These types of differences are shown in the Measure Performance primary category. In some cases, the Impact Evaluation Team found material changes to project performance that had occurred after the Program M&V period. While Program M&V was found to increase the quality of the reported savings, it is still a finite tool. The most significant differences in this category came from projects that received Program M&V and whose loads continued to change thereafter. The most common differences under Measure Performance include changes to the load profile or operating hours that affected energy savings of the project. These types of differences are the least predictable, and least attributable to error on the part of the applicant or Program. Potential improvements targeting these differences are few in number and challenging to implement.

2.2.3 Supercomputer Findings

Supercomputer sites trend toward very large savings, representing 40% of the sampled energy savings in this study. A key finding related to these projects is related to investigation of the baseline efficiency for a given supercomputer. All cases were partially or entirely capacity expansion projects, so the baseline case includes a theoretical “standard efficiency” supercomputer operating at the same loads. The program established a protocol for calculating the baseline efficiency of supercomputers in 2013 (and later revised slightly in 2018), which was thoroughly reviewed as a part of this evaluation. This baseline determination document is available by request.

2.3 Recommendations

- **Calculate and track demand impact in accordance with the New York State Technical Resource Manual.** Going forward, it is recommended that all IPE projects with a component of electrical energy savings have the peak demand impacts calculated in accordance with the New York Technical Resource Manual (NYTM). This will allow the Program to track demand impact values that have been calculated in a uniform manner and within the guidelines of the Department of Public Service and to claim these values in regulatory reporting. Even if not a key metric right now, demand and energy historically have cycled back and forth in terms of relative importance. Grid resiliency and related concerns are gaining visibility. It may be prudent and worth the relatively small marginal effort to estimate demand savings at the same time and with similar rigor as energy savings.
- **Leverage all available site-specific data during the EA phase.** Impactful differences were associated with the EA phase of project review. These differences ranged from the use of non-site-specific data, a misuse or non-use of trend data, and errors in calculations. It is recommended that Technical Reviewers leverage all available site-specific data and review their assumptions with the site to ensure their understanding of the project is in-line with the participant's intent. Many of the differences were preventable, and this recommendation should not incur additional cost to the Technical Reviewers or the Program.
- **Continue with Program M&V and baseline characterization procedures.** As presented in Table 2-7, above, the IPE program has achieved strong realization rates for both electric and natural gas savings for the past two evaluations. These results are largely attributable to the rigorous M&V requirements of the Program, and to the standardized and detailed methodology[8] that is used to characterize the baseline alternative considered in the individual project savings calculations. The Program staff is experienced and successfully leverage these tools, as evidenced by the strong RRs. The Impact Evaluation Team recommends that the IPE program continue with Program M&V and the use of a standardized baseline characterization protocol, as they represent best practices in the implementation of an industrial program, particularly one that considers a wide variety of large and complex custom measures.

[8] Available by request from NYSERDA.

3 Methods

This section describes the methods used to evaluate the impacts of the IPE program. The methods used to estimate the savings are discussed in the following sections.

3.1 Program Data Collection and Project Sampling

This section details the project sampling strategy and data collection efforts for each project. The period of evaluation spanned two separate chapters of program funding: the Energy Efficiency Portfolio Standard Phase 2 (EEPS-2) and the Clean Energy Fund (CEF) Transition. Projects from these two funding chapters were separated and sampled independently.

It is important to note that, for this evaluation, NYSERDA selected two evaluation contractors to co-evaluate the IPE program. ERS was selected as the principal consultant, and ADM Associates served as the supplementary consultant to evaluate any projects that ERS may have had a conflict of interest due to their participation in the IPE program as Outreach Contractor, Technical Reviewer, or Project Manager. Throughout the sample design process, ADM and ERS collaborated to ensure that the sample was robust and the process was transparent to all parties. Once stratified (as discussed below and in Appendix A), projects were randomly sampled according to the design targets. At this point, each sampled project was assessed by NYSERDA, ERS, and ADM to determine whether a conflict of interest existed for ERS. All sampled projects for which a real or perceived conflict was determined to exist were assigned to ADM to evaluate; this included a total of 27 projects. The remaining projects were evaluated by ERS.

3.1.1 Energy Efficiency Portfolio Standard Phase 2-Funded Projects

Stratified ratio estimation (SRE) was used for the sample design for its efficient sampling design; SRE was also a good fit due to the lower sample sizes for a targeted level of precision when there is a strong correlation between the program-reported savings and the evaluated savings. As noted, the sample frame constructed includes all projects with at least one measure completed between 1/1/2014 and 12/29/2017.

3.1.1.1 Upper-Level Stratification

The sample design stratified projects by three project types:

- Industrial process efficiency projects
- Data center process efficiency projects
- Industrial and data center non-process efficiency projects

Sample sizes within each upper-level stratification category were calculated by establishing a confidence/precision target for each one.

The error ratios for the sample design were informed by the results of the previous Program evaluation. Table 3-1 provides the confidence/precision targets and assumed error ratios by stratification category.

Table 3-1. Populations and Target Samples by Upper-Level Strata for EEPS-2 and CEF Transition Funded Projects

	Total		Electric			Natural Gas	
Metric	**Total MMBtu**	**Total kWh**	**Industrial Processes**	**Data Center Processes**	**Industrial and Data Center Non-Processes**	**Industrial Processes**	**Industrial and Data Center Non-Processes**
Sample size	30	55	17	16	22	16	14
Population size	55	188	39	29	135	27	28
Target relative precision	10%	10%	10%	10%	10%	10%	10%
Target relative precision	0.5	0.3	0.3	0.3	0.3	0.5	0.5
% of total reported savings	100%	100%	9%	13%	77%	67%	33%

3.1.1.2 Lower-Level Stratification

The lower-level stratification variable is project size. Size categories were based on the magnitude of project savings for electric or natural gas savings projects, respectively. Four size categories were defined per upper-level stratification category. Cutoffs were established using the method described in the *2004 California Evaluation Framework*.[9]

For each upper-level stratification category, the project size was defined based on the program-reported electric or natural gas savings. The largest size stratum in each segment is a census stratum (all projects evaluated). Additional strata were defined to allow for random sampling of the medium- and small-sized projects in each upper-level stratification category.

Projects in the lowest size stratum that accounted for less than 2% of the total energy savings for the upper-level stratification categories were not evaluated. There are a small number of these projects, and they account for a very small part of the overall program-reported savings and have little effect on the RR. The RR developed for the sample frame was applied to these smaller projects.

After reviewing the final sample population, it was determined that 20 projects included a potential conflict of interest for ERS and were therefore assigned to ADM to review. ADM was assigned another seven projects at NYSERDA's discretion.

[9] TecMarket Works, et al. *The California Evaluation Framework*. Project Number: K2033910. Prepared for the California Public Utilities Commission and the Project Advisory Group. June, 2004. Pages 327 to 339 and 361 to 384. http://www.cpuc.ca.gov/NR/rdonlyres/F14E59AF-25B9-45CE-8B3C-D010C761BE8D/0/CAEvaluationFramework.pdf

Table 3-2. Summary of Project Allocation by Evaluation Firm

Strata		ERS	ADM	Total
Electric	Industrial Process	11	6	17
	Data Center Process	12	4	16
	Industrial & Data Center Non-Process	14	8	22
Gas	Industrial Process	9	7	16
	Industrial & Data Center Non-Process	12	2	14
Total		**58**	**27**	**85**

3.1.2 Clean Energy Fund Transition Projects

A rolling sample methodology was implemented to select the CEF Transition funded projects for evaluation. During the course of the evaluation, four CEF Transition funded projects were identified as ready for evaluation and were included. As all evaluation-ready CEF Transition funded projects were included, this sample is a census.

3.2 Site-Level Measurement and Verification

This section details the site-level M&V activities.

3.2.1 M&V Planning, Outreach, and Data Collection

For each evaluated project, the Impact Evaluation Team calculated the evaluated gross savings based on an independent analysis of the installed measures.

The Impact Evaluation Team gathered documents for each project from the IPE program, relevant NYSERDA program managers, and the technical reviewer firms, such that all existing data could be reviewed and leveraged for evaluation efforts.

After reviewing the available data and assessing the level of rigor required for the site, a Data Collection Instrument was developed that detailed the proposed evaluation approach for the site. The data collection instrument outlined the measures and proposed a level of rigor and savings calculation methodology, any in-situ measurements required, and areas of uncertainty in the approach.

In parallel with the Data Collection Instruments, outreach to site staff at the project locations began. In most cases, the Impact Evaluation Team sought permission to access the facility and review the measure with site staff. They also set up metering equipment or collected trend data.

3.2.1.1 Level of Rigor

A flowchart for determining the IPMVP category level of rigor for each sample can be found in Appendix B. Site verification and metering was primarily based on this guideline. The Impact Evaluation Team sought to leverage existing Program or site-supplied data to limit the cost and on-site interference associated with new data collection. Those projects having high or uncertain variability from the expected savings were subject to increased levels of rigor based on the site-specific needs of the project, which may have included on-site verification; gathering additional metering, trending, or billing data; or making a change to the analysis methodology where necessary.

3.2.2 Project-Level Analysis

After collecting all necessary data, the evaluation engineers produced an independent savings analysis. Efforts were made to duplicate the methodology of the submitted analysis where appropriate, so as not to introduce superfluous differences. Analyses approaches were based on IPMVP recommendations, and Table 3-3 shows the percentage of each project that follows each given pathway.

Table 3-3. Project Analyses by IPMVP Option

IPMVP Option	Percentage of Projects
Basic Verification	0%
Option A or B: Retrofit Isolation	89%
Option C: Whole Facility Regression	8%
Option D: Calibrated Simulation	2%
Total	**100%**

3.2.2.1 NYSERDA Review Process

Site reports were submitted to NYSERDA on a rolling basis. All site reports were individually reviewed by NYSERDA evaluation staff prior to approval.

Critical projects (those with high weighted-impact to the overall evaluation or very low realization rates) were presented by the Impact Evaluation Team on recurring phone calls with NYSERDA evaluation and program staff, to discuss the key differences and potential uncertainties in the site analysis. These discussions continued until all parties understood the result, and both NYSERDA evaluation staff and the evaluation contractor were in agreement about the findings. This review process provided both evaluation and program staff early insight to the findings of the evaluation before aggregate-level findings could be prepared; this also acted as an additional layer of quality assurance review for the site-level findings in cases where issues were identified.

3.3 Aggregate and Differences Analysis

After all site-level gross realization rates were finalized, the Impact Evaluation Team processed the results in an aggregation model. Measure-level results from all sampled projects are fed into the model, and each individual result is multiplied by its project sample weight and is rolled-up to the portfolio level to produce the evaluated Program savings.

Separately, a differences aggregation model was constructed to collect all the quantified differences of the sampled projects. These differences are categorized into three primary discrepancy categories, which refer to the time at which the discrepancy originated:

- Application Review – Differences that occurred during project's initial application phase, which theoretically could have been corrected at the time of the EA.
- Project Installation – Differences that occurred during the project's installation or initial program M&V phase, which theoretically could have been corrected at the time of PIR or M&V.
- Measure Performance – Differences that occurred after all interactions with the program were completed, which would not have been possible for the program to foresee.

Under each primary category, secondary categories delineate the nature of the difference being documented. These secondary categories can be seen in Tables 2-8 and 2-9.