Impact Evaluation of 2010 Custom Gas Installations

Massachusetts Energy Efficiency Programs' Large Commercial & Industrial Evaluation

Prepared for: Massachusetts Energy Efficiency Program Administrators
Massachusetts Energy Efficiency Advisory Council

Middletown, Connecticut, July 6, 2012

Table of Contents

List of Exhibits:

1. Executive Summary

This document summarizes the work performed by the KEMA team, led by ERS, to quantify the actual energy savings due to the installation of Custom Gas measures installed through the Massachusetts Energy Efficiency Program Administrator's (PAs) Commercial and Industrial (C&I) Lost Opportunity and Large Retrofit programs in 2010.

This is the second state-wide evaluation of the large C&I custom gas programs in Massachusetts. The primary mission of the study was to determine program realization rates. The rates will be used for planning and program reporting, including program year 2011 annual reporting and any 2013-2015 program planning and subsequent year reporting, unless replaced by results from a subsequent study. The gas program PAs include: Columbia Gas, National Grid Gas, NSTAR Gas, Berkshire Gas, New England Gas and Unitil. This evaluation effort received oversight by the Energy Efficiency Advisory Council (EEAC) consultant representative.

The scope of work for this impact evaluation included all the 2010 Custom Gas measures including high efficiency heating equipment, heating systems, heating controls (e.g., energy management systems [EMS]), boiler combustion controls, building shell measures, high efficiency gas industrial process equipment, and other measures.

1.1 Methods

The evaluation realization rate results were derived from on-site engineering monitoring- based assessments of installed custom gas measures of a statistically representative sample of forty-eight sites.

1.1.1 Sampling Strategy

Many potential sample designs were considered in an attempt to achieve the project goals of 80% confidence and ±10% relative precision overall. In the final design, which included 48 sites, the evaluation team expected to achieve this statewide goal, as well as about ±20% relative precision for Columbia Gas, National Grid and NSTAR. Table 1-1 shows the distribution and estimated precision of sample sites in this design.

Table 1-1: Estimated Precision for Final Sample Design

Program Administrator	Projects	Total Therms Saved	Error Ratio	Confidence Level	Planned Sample Size	Anticipated Relative Precision
Berkshire Gas	8	89,684	0.7	80%	2	±24.66%
Columbia Gas	95	1,553,740	0.7	80%	13	±16.34%
New England Gas	1	23,400	0.7	80%	1	±0.00%
National Grid	109	1,710,500	0.7	80%	17	±18.44%
NSTAR	53	938,625	0.7	80%	13	±18.66%
UNITIL	6	111,412	0.7	80%	2	±34.75%
Total	272	4,427,361	0.7	80%	48	±10.01%

This sample design excluded sixty-three sites with site savings of less than a thousand therms. These sites accounted for less than 1% of the total tracking savings and were subject to a project file review of a sample of those sites.

1.1.2 Measure Savings Estimates

Following the final sample selection of 2010 Custom Gas applications and prior to beginning any site visits, the engineers acquired all available information concerning the measure at the site including project files and electronic analysis, supplemented with interviews of PA staff, customers, and vendors involved with implementation of the measures. A particular effort was made to gather gas billing impacted by the project. Detailed measurement and evaluation plans were produced for each site and reviewed by PA study managers in most cases, except for a few of the replacement sites, where timing was critical, towards the end of the study period.

Each site was visited by a lead engineer and a junior engineer. Activities included visual inspection of the installed measures, acquisition of nameplate data, spot measurements of boiler efficiencies, interviews with knowledgeable site staff, review of plans, and placement of logger equipment. Depending upon the measure under evaluation, loggers were placed to measure parameters such as supply air temperatures, return and supply water temperatures, and motor runtime profiles. The study design provided for a longer than normal two week logging period. Loggers were left in place for a minimum of a week and for as much as eight weeks. When possible, trend data was secured from the building automation system. Capturing distribution company billing data for all the affected meters at a site was a goal as well.

The engineer selected from a variety of analytical techniques as appropriate for the measure including eQuest building simulation models and bin temperature models. Customer gas bills were used to calibrate bills, corroborate savings, and in some cases used as the primary means of determining savings impacts.

Finally, a sample of 20 sites was selected at random from the 63 sites with site savings less than a 1000 therms. The review includes an assessment of the information available in the files, the savings calculation methods, and also a report of the measure savings fraction as a percent of the bills.

1.2 Findings and Results

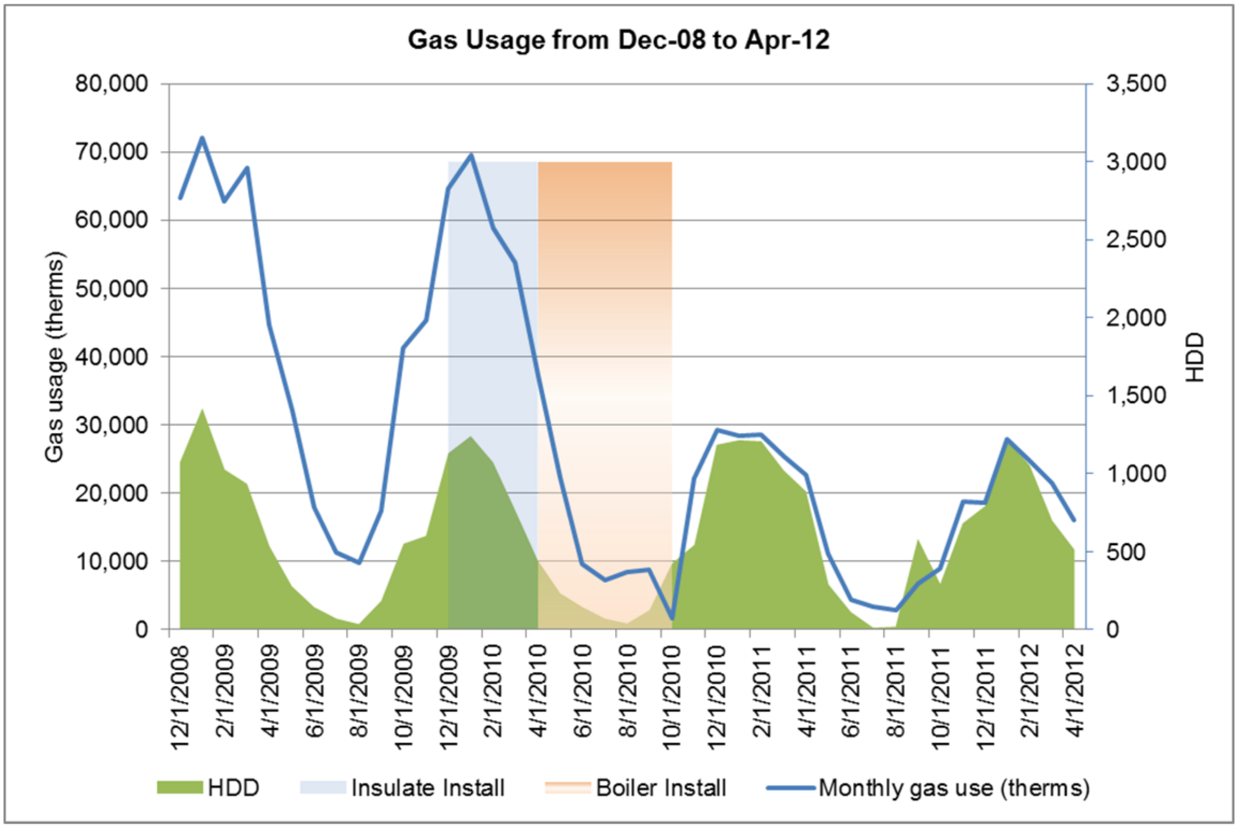
The program administrators are to be commended for the significant increases in program participation and in the breadth of measures implemented when comparing the 2009 and 2010 program years. Tracking savings increased from 2.0 to 4.4 million therms, while the number of unique accounts increased from 249 to 330 from 2009 to 2010, respectively. More impressive is the range and complexity of projects. The 2010 program year included:

- Regenerative thermal oxidizer measures which provide a high level of heat recovery when incinerating volatile organic compounds (VOCs) in industrial processes;
- Expansion of two production lines, one manufacturing a food product, the second in a coating process for large production runs of sheet metal housing units; and
- Ingenious heat recovery systems, including the capture of waste heat in a steam tunnel and capturing of an air compressor rejected heat for space heating;
- Multi-measure public housing projects replacing, in one case, 191 boilers and DHW systems.

The PA's are successfully reaching a broader range of customers, particularly in the industrial sector. In the 2009 the sample population, measures directly affecting a process accounted for a single project at one of the four manufacturing/ industrial sites, while in 2010, industrial process measures accounted for five of the projects at thirteen manufacturing/industrial sites.

These projects often provided dramatic reductions in customer bills which will persist over decades. For example, at Site A, a public housing complex, the bills were reduced by 64%, as shown in Figure 1-1 with a combination of prescriptive and custom gas measures.

Figure 1-1: Site A Bill Impact of Energy Efficiency Project



Although 2010 showed an expansion in customers, types of measures, and tracking savings, the final statewide realization rate of 67.7% was disappointing. The previous year's evaluation yielded an 88% realization rate with an outlier site, and 71% without that site factored in. There are challenges when doubling production and including a much broader range of measures with unique base lines and unique technical assessment. The results of the evaluation imply administrative growing pains. Administrative errors and limited technical back-up of estimates contributed to variances in realization rates. The PAs are taking steps to improve the technical review. However, the realization rates between sites approved in 2010 and approved prior to 2010 were almost identical. This might be expected considering that the recommendations from the last evaluation were not available until mid-2011.

In conducting the evaluation, the sampling goals were met by PA and by strata. In preparation for analyzing the evaluation results collected for the Custom Gas sample points, the original 2010 population stratum boundaries were used to calculate case weights for each sample observation. The site-level evaluation results were aggregated. The statewide results are summarized in Table 1-2.

Table 1-2: Summary of Custom Gas Results

Statistic	Annual Therms
All Program Administrators	
Total Tracking Savings	4,427,361
Total Measured Savings	2,991,776
Realization Rate	67.6%
Relative Precision at 80% Confidence	±9.0%
Error Bound at 80% Confidence	268,703
Sample Size	48
Error Ratio	0.50

The results by PA are presented in Table 1-3.

Table 1-3: Summary of Results by PA

PA	Annual Therms	PA	Annual Therms
Berkshire Gas		**National Grid**	
Total Tracking Savings	89,684	Total Tracking Savings	1,710,500
Total Measured Savings	34,660	Total Measured Savings	1,172,176
Realization Rate	38.6%	Realization Rate	68.5%
Relative Precision at 80% Confidence	±0.8%	Relative Precision at 80% Confidence	±17.4%
Error Bound at 80% Confidence	276	Error Bound at 80% Confidence	203,593
Sample Size	2	Sample Size	17
Error Ratio	0.02	Error Ratio	0.62
Columbia Gas		**NSTAR**	
Total Tracking Savings	1,553,740	Total Tracking Savings	938,625
Total Measured Savings	1,293,037	Total Measured Savings	444,200
Realization Rate	83.2%	Realization Rate	47.3%
Relative Precision at 80% Confidence	±12.9%	Relative Precision at 80% Confidence	±11.2%
Error Bound at 80% Confidence	167,329	Error Bound at 80% Confidence	49,693
Sample Size	13	Sample Size	13
Error Ratio	0.42	Error Ratio	0.39
New England Gas		**Unitil**	
Total Tracking Savings	23,400	Total Tracking Savings	111,412
Total Measured Savings	12,902	Total Measured Savings	34,801
Realization Rate	55.1%	Realization Rate	31.2%
Relative Precision at 80% Confidence	±0.0%	Relative Precision at 80% Confidence	±48.3%
Error Bound at 80% Confidence	-	Error Bound at 80% Confidence	16,807
Sample Size	1	Sample Size	2
Error Ratio	0.00	Error Ratio	0.86

Table 1-4 presents results by measure category.

Table 1-4: Results by Measure Category

Measure	Annual Therms	Measure	Annual Therms
Controls		**Non-Boiler Heating**	
Total Tracking Savings	1,294,158	Total Tracking Savings	599,531
Total Measured Savings	828,761	Total Measured Savings	347,588
Realization Rate	64.0%	Realization Rate	58.0%
Relative Precision at 80% Confidence	±18.5%	Relative Precision at 80% Confidence	±20.3%
Error Bound at 80% Confidence	153,043	Error Bound at 80% Confidence	70,547
Sample Size	21	Sample Size	12
Error Ratio	0.75	Error Ratio	0.69
Envelope		**Other**	
Total Tracking Savings	598,368	Total Tracking Savings	949,780
Total Measured Savings	548,562	Total Measured Savings	618,864
Realization Rate	91.7%	Realization Rate	65.2%
Relative Precision at 80% Confidence	±26.3%	Relative Precision at 80% Confidence	±14.2%
Error Bound at 80% Confidence	144,359	Error Bound at 80% Confidence	87,853
Sample Size	8	Sample Size	13
Error Ratio	0.55	Error Ratio	0.67
Hydronic/Steam			
Total Tracking Savings	985,524		
Total Measured Savings	650,981		
Realization Rate	66.1%		
Relative Precision at 80% Confidence	±6.7%		
Error Bound at 80% Confidence	43,712		
Sample Size	14		
Error Ratio	0.43		

As part of the engineering analysis, each site was reviewed to identify factors that created discrepancies between evaluated and tracked savings and then to categorize and quantify them. The intent of this analysis is to provide the PAs with indicators of where to focus future process improvements.

1.3 Recommendations

Overall, the Custom Gas program is successfully providing energy savings in the State of Massachusetts. A summary of the recommendations follows.

1.3.1 Realization Rates

The study produced statewide results that are reliable (±9.0%) at 80% confidence. In addition, the precision levels for several PAs (National Grid: ±17.4%, Columbia Gas: ±12.9%, and NSTAR: ±11.2%) are sufficient to warrant application of the three individual PA realization results to the 2011 results according to the protocol established in the November 2010 Protocol memo, which stated that individual realization rates may be applied for those PAs with more than ten sites and a final precision better than ±9.0%.

The results do not support application by measure, although on first glance, the statistical outcomes may indicate otherwise. The measure classifications were defined for the purposes of the evaluation and are not necessarily consistently applied in tracking. These results indicate performance trends, but are not considered reliable for planning.

1.3.2 Program Implementation Improvements

The evaluation team reviewed project files, conducted detailed analysis of the information provided in the files, and quantified discrepancies analysis to make the recommendations of this section. The recommendations are in summary:

- Project documentation should include savings estimates in the native file form and support the claimed baseline.
- The baselines for HVAC equipment replacement projects will be building code in the large majority of cases. The PA's need to document alternate baselines.
- Boiler control measures will typically save between 1-5% in gas usage. Projected savings larger than 5% should be scrutinized closely. The baseline for a fuel switching burner with controls is based on the building code efficiencies. Applications should include measurement of the pre-installation combustion efficiency across multiple firing ranges.
- High temperature, 100% outdoor air, direct fired units used as heating units should demonstrate a need for the high levels of ventilation. Units in excess of 5000 CFM require heat recovery according to the energy code.
- Five vendors in the regenerative thermal oxidizer market were surveyed; the results were used to define a lowest cost, reasonable option baseline for this technology.
- PA's should consider requiring a commissioning phase for unique and complex process measures.

1.3.3 Evaluation recommendations

The evaluation team recommends considering improvements to the evaluation process as follows:

- Begin a gas evaluation as early as possible to ensure metering of a range of winter temperatures;
- Consider having the evaluator review and select project files for copies on the PA premises;
- Billing data has proven to be invaluable. PA and evaluators should work together to ensure a full mapping of accounts serving a measure.

Two evaluations have been completed in quick succession: the 2009 program year evaluation with results reported in mid-2011 and this 2010 program year evaluation with results in mid-2012. Since the 2011 program year was well underway, the 2009 evaluation recommendations are expected to have minimal impact on 2011 projects. It should be noted, however, that the PAs had been pursing improvements to program procedures independent of evaluation recommendations.

Although the prior evaluation recommendations were too late to substantially impact the 2011 program year, the PAs and EEAC consultants may still wish to consider evaluating the 2011 program. It is possible that the process improvements put into place by the PAs will improve realization rates and there is also much that can be learned from an evaluation in terms of program policy and technology performance. As an intermediate step, the PAs and EEAC consultants may wish to consider an initial evaluation test which would fall short of actual site work. In this test, the evaluators would conduct a sample design and gather paperwork and billing data to ascertain the quality of the PA technical review, the disposition of measure types, baseline issues, and also conduct an initial billing analysis. The results of this process can be used to decide whether sufficient progress has been made to proceed to full M&V activities or postpone a study until the following year.

As a final recommendation, the evaluators recommend measuring code compliant single or two-stage boilers to develop an improved baseline for boiler measures which would characterize the field performance of a standard boiler.

2. Description of Sampling Strategy

The goal of this sample design was to evaluate sites to produce aggregated realization rates for Custom Gas projects by PA with reasonable precision, given a project budget that would support about 40 monitored sites. Several alternative approaches and designs were evaluated, but given the decision to evaluate all measures at each selected site, designs based on measure categories were rejected. The recommended strategy developed by KEMA and ERS is to select the target number of sites by PA, and then evaluate all of the energy efficiency measures encountered at those sites.

The population frame for this impact evaluation is the set of custom gas projects rebated in 2010, as tracking system data provided by the six PAs in Massachusetts. Sites with annual savings less than 1,000 therms were deleted, leaving 272 sites with a total savings of 4,427,361 therms. Table 2-1 shows the distribution of all tracking system records, based on annual savings in therms, by PA. The largest savings are attributable to projects implemented by Columbia Gas and National Grid, with NSTAR also having a substantial share.

Table 2-1: Population Summary Statistics

Program Administrator	Projects	Total Therms	Average Therms	Minimum	Maximum	StdDev	CV
Berkshire Gas	8	89,684	11,211	1,000	75,188	24,195	2.16
Columbia Gas	95	1,553,740	16,355	1,012	288,449	45,170	2.76
New England Gas	1	23,400	23,400	23,400	23,400	0	0.00
National Grid	109	1,710,500	15,693	1,099	152,674	26,407	1.68
NSTAR	53	938,625	17,710	1,028	134,992	25,636	1.45
UNITIL	6	111,412	18,569	1,150	79,715	27,540	1.48
Total	272	4,427,361					

2.1 Sample Design

The parameters considered in the sample design are the number of sample observations planned and the anticipated error ratio of quantity being estimated. The error ratio is a measure of the strength of the relationship between the known characteristic (e.g., tracking system savings) and the quantity being estimated (e.g., evaluated savings). Based on the results of the 2009 Custom Gas Impact Evaluation, the error ratio was calculated to be 0.68. For planning this study, we elected to use a reasonable conservative error ratio of 0.7 for the 2010 sample design.

Many potential sample designs were considered in an attempt to achieve the project goals of 80% confidence and ±10% relative precision overall. In the final design, which included 48 sites, the evaluation team expected to achieve this statewide goal, as well as about ±20% relative precision for Columbia Gas, National Grid and NSTAR. Table 2-2 shows the stratum cut points and distribution of sample sites in this design.

Table 2-2: Final Sample Design

Program Administrator	Stratum	Maximum Therms Saved	Number of Sites	Total Therms Saved	Sample Size	Inclusion Probability
Berkshire Gas	1	3,357	7	14,496	1	0.1429
Berkshire Gas	2	75,188	1	75,188	1	1.0000
Columbia Gas	1	4,829	64	162,727	3	0.0469
Columbia Gas	2	27,184	20	203,286	3	0.1500
Columbia Gas	3	77,520	7	329,871	3	0.4286
Columbia Gas	4	288,449	4	857,856	4	1.0000
New England Gas	1	23,400	1	23,400	1	1.0000
National Grid	1	9,947	70	241,519	4	0.0571
National Grid	2	24,696	20	334,295	4	0.2000
National Grid	3	46,451	10	349,085	4	0.4000
National Grid	4	131,953	7	490,068	3	0.4286
National Grid	5	152,674	2	295,533	2	1.0000
NSTAR	1	13,090	35	176,886	4	0.1143
NSTAR	2	31,970	10	234,493	4	0.4000
NSTAR	3	68,544	6	281,362	3	0.5000
NSTAR	4	134,992	2	245,884	2	1.0000
UNITIL	1	11,680	5	31,697	1	0.2000
UNITIL	2	79,715	1	79,715	1	1.0000

Table 2-3 lists the calculated precision estimates for this design, following stratification.

Table 2-3: Estimated Precision for Final Sample Design

Program Administrator	Projects	Total Therms Saved	Error Ratio	Confidence Level	Planned Sample Size	Anticipated Relative Precision
Berkshire Gas	8	89,684	0.7	80%	2	±24.66%
Columbia Gas	95	1,553,740	0.7	80%	13	±16.34%
New England Gas	1	23,400	0.7	80%	1	±0.00%
National Grid	109	1,710,500	0.7	80%	17	±18.44%
NSTAR	53	938,625	0.7	80%	13	±18.66%
UNITIL	6	111,412	0.7	80%	2	±34.75%
Total	272	4,427,361	0.7	80%	48	±10.01%

2.2 Final Sample

Table 2-4 presents the final sample disposition. The sample quotas were fulfilled by PA and by strata. Of the 48 sites, 39 were primary samples, while the balance of sites required recruiting one or more back-up site. In the course of recruiting for these 9 sites, 18 additional sites were considered and ultimately dropped.

Table 2-4: Final Sample Disposition

Program Administrator	Original Sample	Sites w Back-ups	Attempted Recruits	Final
Columbia Gas	13	3	7	13
Berkshire Gas	2	0	0	2
New England Gas	1	0	0	1
National Grid	17	4	14	17
NSTAR	13	1	4	13
UNITIL	2	1	2	2
Total	48	9	27	48

The reasons for the drops are summarized in Table 2-5. Of the dropped sites, one of the sites was determined to be a CHP project and three of the sites had prescriptive measure only. The remaining 14 sites were dropped because the customer would not respond to recruitment efforts or refused the site visit.

Table 2-5: Reasons for Dropped Sites

Program Administrator	Prescriptive or CHP	Site Refused
Columbia Gas	1	3
Berkshire Gas		
New England Gas		
National Grid	3	7
NSTAR		3
UNITIL		1
Total	4	14

Table 2-6 summarizes the final sites for which monitoring and verification activities were completed and indicates the PA, the number of measures at the site, the application type (R for retrofit, LO for Lost opportunity), the verified measure description, and finally, the facility type.

Table 2-6: Strata by Site ID for Final Selection

Site	PA	Numb of Mea	App Type	Strat-um	Verified Measures	Facility Type
011	BERKSHIRE	1	LO	1	Boiler replacement	Town hall
018	BERKSHIRE	2	R	2	EMS	Manufacturing facility
065	COLUMBIA	1	LO	1	Gas-fired rooftop units	Warehouse
095	COLUMBIA	2	LO	1	Boiler and indirect DHW	Apartments
101	COLUMBIA	1	R	1	Envelope roof insulation	Museum archival
109	COLUMBIA	1	LO	2	Condensing boiler replacement, space heating, and DHW	Temple
121	COLUMBIA	1	R	2	Dock seals	Manufacturing facility
122	COLUMBIA	2	LO	2	Condensing boilers replacement, indirect DHW	LI Multi-family housing
125	COLUMBIA	1	R	3	Ventilation heat recovery	Manufacturing facility
128	COLUMBIA	3	R	3	Envelope, windows, boiler replacement	Hospital
130	COLUMBIA	5	LO and R	3	Boiler replacement and non-boiler heating. Decentralized boiler.	Hospital
132	COLUMBIA	2	LO and R	4	Condensing Boiler replacement, hood control for RTO measure	Manufacturing facility
133	COLUMBIA	1	R	4	Heat recovery with RTO	Manufacturing facility
134	COLUMBIA	4	R	4	EMS, condensate tank insulation, non-boiler heating replacement, process, renewable gas	Sewage treatment plant
135	COLUMBIA	2	R	4	Boiler controls, boiler other	Manufacturing facility
136	NEGAS	1	R	1	Process Catalytic oven and other process	Manufacturing facility
185	NGRID	2	R	1	Boiler steam replacement and heat timer	Multifamily
207	NGRID	2	R	1	Kitchen hoods and heat recovery	Restaurant
219	NGRID	1	R	1	EMS	Church facility
229	NGRID	1	R	1	Direct fired space heating	Warehouse
235	NGRID	1	R	2	Heat pipe heat recovery	Medical research and lab
238	NGRID	1	R	2	Process improvements	Food manufacturing
242	NGRID	1	R	2	Controls, programmable thermostats	University dormitory
250	NGRID	2	R	2	Envelope roof insulation, direct fired heating replacement.	Manufacturing facility
253	NGRID	2	R	3	EMS and pipe insulation	High School
257	NGRID	1	R	3	Boiler controls	University boiler plant

Site	PA	Numb of Mea	App Type	Strat-um	Verified Measures	Facility Type
258	NGRID	1	R	3	EMS	Middle School
259	NGRID	1	R	3	Process heat recovery	Food manufacturing
264	NGRID	1	R	4	Vent constant to VAV, controls	Office and lab space
267	NGRID	3	R	4	Blow down optimization, boiler controls, other.	Medical clinic and offices
268	NGRID	2	R	4	Envelope attic and wall	LI Multi-family housing
269	NGRID	1	R	5	RTO	Manufacturing facility
270	NGRID	4	LO and R	5	Thermostats, wall insulation, windows, replacement boiler	LI Multi-family housing
275	NSTAR	1	R	1	EMS	Retail store
282	NSTAR	1	R	1	TRV	University dormitory
286	NSTAR	1	R	1	TRV	University dormitory
306	NSTAR	1	R	1	Boiler replacement, condensing	Nursing home
313	NSTAR	2	LO	2	Boiler condensing and DHW replacement	University dormitory
315	NSTAR	1	R	2	Boiler controls	Hospital
317	NSTAR	1	R	2	Boiler controls	High School
318	NSTAR	2	R	2	Solar thermal DHW and heat recovery	University dormitory
321	NSTAR	1	R	3	EMS in tracking, but boiler controls O2 trim.	Medical research and lab
323	NSTAR	3	R	3	Direct fired units, insulation and EMS	Manufacturing facility
325	NSTAR	2	R	3	Boiler replacement with condensing and indirect DHW	LI Multi-family housing
326	NSTAR	2	R	4	Boiler replacement with condensing and heat recovery	Medical research and lab
327	NSTAR	2	LO	4	Boiler replacement with condensing, indirect DHW	LI Multi-family housing
332	UNITIL	1	R	1	Roof insulation	Community hall
335	UNITIL	1	R	2	Boiler controls	Medical clinic and offices

3. Description of Methodology

This section describes the site methodology generally for both the development of site evaluation plans, the execution of the plans, and the final process for producing program results.

3.1 Measurement and Evaluation Plans

Following the final sample selection of 2010 Custom Gas applications and prior to beginning a site visit, ERS developed detailed measurement and evaluation plans applications. These plans outlined on-site methods, strategies, monitoring equipment placement, calibration, and analysis issues. The PAs provided comments and edits to clarify and improve the plans prior to them being finalized.

Evaluators utilized the savings analysis methodologies from the Technical Assistance study (TA) whenever possible. However, in many cases, the TA methodology was unavailable or found to be incorrect or inappropriate. In those cases, the evaluators performed an analysis more appropriate to the measure being evaluated. Adjustments to savings methodologies were presented and agreed upon in the measurement and evaluation plans.

The site evaluation plan played an important role in establishing approved field methods and ensuring that the ultimate objectives were met.

3.2 On-Site Data Gathering, Analysis, and Reporting

Data collection included physical inspection and inventory, interview with facility personnel, observation of site operating conditions and equipment, and short-term metering. At each site, the evaluator performed a facility walk-through that focused on verifying the post-retrofit or installed conditions of the energy efficiency measure. Several of the facilities utilized EMS controls which were either part of the application itself or controlled equipment that was included in the application. Evaluators viewed EMS screens to verify schedules and operating parameters where applicable. At times, the EMS was utilized to log key parameters, or previously trended data was extracted from the system. Instrumentation such as current loggers, motor status, and temperature loggers were installed to monitor the usage of the installed HVAC equipment and associated affected spaces. At most sites, combustion efficiency measurements were taken of the heating equipment. Gas bills were acquired from the gas distribution company and from customer records.

Weather sensitive measures were assessed using historical weather data from periods matching the metering period or the gas billing. Savings estimates were normalized to a typical year using a typical meteorological year (TMY3). Weather stations located closest to each facility were used for all weather-sensitive calculations.

Each site report details the analysis methods used specific to each project including algorithms, assumptions, and calibration methods where applicable. The actual analytical techniques employed depended upon the applicant's methods, the measure, and site conditions. The methods included:

> **Hourly temperature spreadsheet models** Most condensing boiler, boiler, boiler controls, EMS, heat recovery, and water heater savings were estimated using an 8760 hour model. Historical hourly weather data for a twelve month post installation period forms the basis of the model, permitting an hourly calculation of thermal load and equipment efficiency. The temperature and runtime logged measurements are utilized to identify a relationship between operation and outdoor air temperature. Operating schedules are also incorporated into the model. Boiler efficiency is based on the measured efficiencies extrapolated across the firing range of the boiler. For condensing boilers, the latent efficiency component was typically modeled as a function of the return water temperature. The final model is usually calibrated to actual customer bills.
>
> **Bin temperature spreadsheet models** A bin temperature model is a simplified version of the hourly model. While the thermal load and efficiency calculations are similar, the weather is represented by the number hours of occurrence of an outdoor temperature by temperature bin (usually in five degree increments).The bin model was used in cases where the applicant had also used a bin model and for some of the simpler measures.
>
> **Building simulation models** Most of the envelope measures including attic insulation, roof insulation, wall insulation, and window replacement were generally modeled using a simple eQUEST building simulation model. The building simulation model captures impacts of thermal mass and solar gains, which can be important for envelope measures. One of the most complex sites, a new construction project incorporating enhanced insulation and advanced HVAC design with radiant heat and cold beam cooling was modeled with an eQUEST model. The building models incorporated field measurements and observations, such as boiler efficiency measurements and building schedules. Models were generally calibrated to customer monthly gas bills.

Billing analysis A few sites, like the temperature-controlled radiator valve (TRV) sites, were evaluated using a two-sided billing analysis, where the savings was determined to the difference between the weather adjusted pre and post billing data. Billing analysis was used if the baseline conditions could not be confirmed and no other significant changes had occurred at the site. In some cases, a one-sided billing analysis was used, where the current facility load was determined from the post-installed weather normalized billing data. The pre and post-efficiency conditions were then applied to the determined gas use to calculate the savings. Bills were reviewed in all cases to ensure the results were reasonable in light of the bills.

At almost all of the sites, customer billing usage was used to corroborate the savings. Engineers submitted draft site reports to the PAs upon completion of each site evaluation, which after review and comment resulted in the final reports. These are included in Appendix B.

3.3 Aggregate Analysis Procedures

In order to aggregate the individual site results from the Custom Gas sample, KEMA applied the model-assisted stratified ratio estimation methodology.[1][2] The key parameter of interest is the population realization rate, i.e., the ratio of the evaluated savings for all population projects divided by the tracking estimates of savings for all population projects. This rate is estimated for the overall Massachusetts program, as well as for individual PAs. Of course, the population realization rate is unknown, but it can be estimated by evaluating the savings in a sample of projects. The sample realization rate is the ratio between the weighted sum of the evaluated savings for the sample projects divided by the weighted sum of the tracking estimates of savings for the same projects. The total tracking savings in the population is multiplied by the sample realization rate to estimate the total evaluated savings in the population. The statistical precisions and error ratios are calculated for each level of aggregation.

3.4 Small Site Savings Project Review

As noted previously, sixty-three sites with less than 1000 therms of annual savings have been screened from the sample. The site measures, per the tracking data, include: thermostats, door

[1] [1] The California Evaluation Framework, prepared for Southern California Edison Company and the California Public Utility Commission, by the TecMarket Works Framework Team, June 2005, Chapters 12-13.

[2] [2] Model Assisted Survey Sampling, C. E. Sarndal, B. Swensson, and J. Wretman, Springer, 1992.

and window replacements, air-sealing, attic and roof insulation, and pipe insulation. A random sample of 20 sites was selected for review and measure characterization, although not for impact assessment. The primary purpose of the review is to determine if this population has been properly categorized and that the savings are reasonable.

Results are presented in Appendix A.

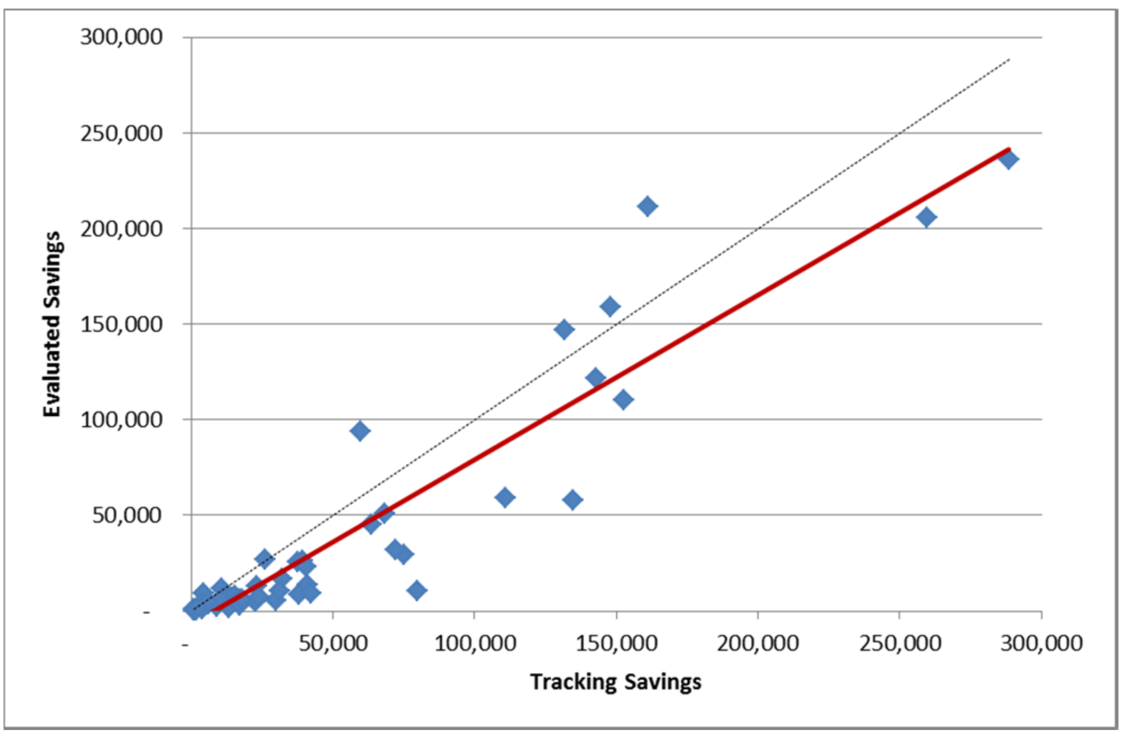
4. Results

This section presents the site and population level results. The site level results include the level estimates of savings and a quantitative breakdown of the factors that caused the realization rates to deviate from 100%. The population level analysis includes a presentation of the final case weights and the resulting realization rates.

4.1 Site Level Results

Figure 4-1 presents a scatter plot of evaluation results for annual therm savings plotted against the PA tracking savings. The dashed line represents a realization rate of one. The slope of the solid line in this graph is an indication of the overall realization rate and how it relates to a realization rate of 100%. These sample data are scattered widely around the trend line, which supports the estimate made during the design process that the error ratio would be relatively high.

Figure 4-1: Scatter Plot of Evaluation Results for Annual Energy Savings



Unlike last year, there are no outliers that unduly influence the outcome.

4.2 Retrospective Realization Rates

In preparation for analyzing the evaluation results collected for the Custom Gas sample points, the original 2010 population stratum boundaries were used to calculate case weights for each sample observation. These weights reflect the number of projects that each of the sample points represent in their respective populations and allow for the aggregation of results across strata and PAs. The final case weights for the study, which reflect sample substitutions, are shown in the last column in Table 4-1.

Table 4-1: Custom Gas Case Weights

Measure	PA	Stratum	Total Projects	Total Annual Therms	Projects in Sample	Case Weight
Custom	BERKSHIRE	1	7	14,496	1	7.00
Custom	BERKSHIRE	2	1	75,188	1	1.00
Custom	COLUMBIA	1	64	162,727	3	21.33
Custom	COLUMBIA	2	20	203,286	3	6.67
Custom	COLUMBIA	3	7	329,871	3	2.33
Custom	COLUMBIA	4	4	857,856	4	1.00
Custom	NEGAS	1	1	23,400	1	1.00
Custom	NGRID	1	70	241,519	4	17.50
Custom	NGRID	2	20	334,295	4	5.00
Custom	NGRID	3	10	349,085	4	2.50
Custom	NGRID	4	7	490,068	3	2.33
Custom	NGRID	5	2	295,533	2	1.00
Custom	NSTAR	1	35	176,886	4	8.75
Custom	NSTAR	2	10	234,493	4	2.50
Custom	NSTAR	3	6	281,362	3	2.00
Custom	NSTAR	4	2	245,884	2	1.00
Custom	UNITIL	1	5	31,697	1	5.00
Custom	UNITIL	2	1	79,715	1	1.00

The site-level evaluation results were aggregated using stratified ratio estimation. The PA realization rates were estimated and then applied to each PA's total tracking savings to determine their total measured savings. The state-wide realization rate is the ratio of the total measured savings to the total tracking savings, each of which is calculated by summing across the PAs. Table 4-2 summarizes the state-wide results of this analysis. The realization rate for Custom Gas measures was found to be 67.6%. This is somewhat lower than the analysis of 2009 projects (88% with the outlier and 71% without the outlier). The relative precision for this estimate was found to be ±9.0% at the 80% level of confidence. The error ratio was found to be 0.50, which is lower than the 0.70 used during the sample design.

Table 4-2: Statewide Results

Statistic	Annual Therms
All Program Administrators	
Total Tracking Savings	4,427,361
Total Measured Savings	2,991,776
Realization Rate	67.6%
Relative Precision at 80% Confidence	±9.0%
Error Bound at 80% Confidence	268,703
Sample Size	48
Error Ratio	0.50

The results of the analysis of realization rates by PA follow in Table 4-3. It was anticipated that National Grid, Columbia Gas and NSTAR would have enough sample points to produce estimates with adequate precision to use their individual results. The results indicate that the goal of at least ±20% for these PAs was achieved.

Table 4-3: Results by Program Administrator

PA	Annual Therms	PA	Annual Therms
Berkshire Gas		**National Grid**	
Total Tracking Savings	89,684	Total Tracking Savings	1,710,500
Total Measured Savings	34,660	Total Measured Savings	1,172,176
Realization Rate	38.6%	Realization Rate	68.5%
Relative Precision at 80% Confidence	±0.8%	Relative Precision at 80% Confidence	±17.4%
Error Bound at 80% Confidence	276	Error Bound at 80% Confidence	203,593
Sample Size	2	Sample Size	17
Error Ratio	0.02	Error Ratio	0.62
Columbia Gas		**NSTAR**	
Total Tracking Savings	1,553,740	Total Tracking Savings	938,625
Total Measured Savings	1,293,037	Total Measured Savings	444,200
Realization Rate	83.2%	Realization Rate	47.3%
Relative Precision at 80% Confidence	±12.9%	Relative Precision at 80% Confidence	±11.2%
Error Bound at 80% Confidence	167,329	Error Bound at 80% Confidence	49,693
Sample Size	13	Sample Size	13
Error Ratio	0.42	Error Ratio	0.39
New England Gas		**Unitil**	
Total Tracking Savings	23,400	Total Tracking Savings	111,412
Total Measured Savings	12,902	Total Measured Savings	34,801
Realization Rate	55.1%	Realization Rate	31.2%
Relative Precision at 80% Confidence	±0.0%	Relative Precision at 80% Confidence	±48.3%
Error Bound at 80% Confidence	-	Error Bound at 80% Confidence	16,807
Sample Size	1	Sample Size	2
Error Ratio	0.00	Error Ratio	0.86

Analyses were also performed by measure category (across PAs). These results are presented in Table 4-4. While the precisions are better than 20% for all but envelope and non-boiler heating measures, the categories were assigned by the evaluators and may not map back to the PA defined measure categories, and therefore should not be applied directly to measure estimtes.

Table 4-4: Results by Measure Category

Measure	Annual Therms	Measure	Annual Therms
Controls		**Non-Boiler Heating**	
Total Tracking Savings	1,294,158	Total Tracking Savings	599,531
Total Measured Savings	828,761	Total Measured Savings	347,588
Realization Rate	64.0%	Realization Rate	58.0%
Relative Precision at 80% Confidence	±18.5%	Relative Precision at 80% Confidence	±20.3%
Error Bound at 80% Confidence	153,043	Error Bound at 80% Confidence	70,547
Sample Size	21	Sample Size	12
Error Ratio	0.75	Error Ratio	0.69
Envelope		**Other**	
Total Tracking Savings	598,368	Total Tracking Savings	949,780
Total Measured Savings	548,562	Total Measured Savings	618,864
Realization Rate	91.7%	Realization Rate	65.2%
Relative Precision at 80% Confidence	±26.3%	Relative Precision at 80% Confidence	±14.2%
Error Bound at 80% Confidence	144,359	Error Bound at 80% Confidence	87,853
Sample Size	8	Sample Size	13
Error Ratio	0.55	Error Ratio	0.67
Hydronic/Steam			
Total Tracking Savings	985,524		
Total Measured Savings	650,981		
Realization Rate	66.1%		
Relative Precision at 80% Confidence	±6.7%		
Error Bound at 80% Confidence	43,712		
Sample Size	14		
Error Ratio	0.43		

The mapping of measures to measure categories is shown in Table 4-5.

Table 4-5: Mapping of Measures to Measure Category

New	Old
Hydronic/Steam Systems replacements including all steam or hot water systems serving space heating, DHW, process or any combination thereof.	**Boilers** Boilers, burner, and controls controls.
Controls Boiler controls, ventilation control, thermostats, TRVs, EMS, steam valve	**Controls** Ventilation control, thermostats, TRVs, EMS
Envelope Roof, wall and floor insulation; windows, destratification fans, infiltration reductions (though not forced ventilation)	**Insulation** Roof, wall and floor insulation; windows,
Non-boiler heating Direct fired, infrared heat, solar thermal, furnace	**Non-boiler heating** Direct fired, infrared heat, solar thermal, furnace. Also DHW
Other Pipe insulation, pool, heat recovery, steam traps, non-boiler replacement process measures.	**Other** Pipe insulation, pool, heat recovery, steam traps, process measures. Also destrat fans, ventilation controls, indirect DHW
	Other heating Equipment replacement, likely process.

4.3 Individual Implementation and Technology Observations

The evaluators observed certain implementation practice and technology trends that are summarized in this section. These observations results in specific recommendations presented in the next section. Each section begins with an illustrative example shown in italics.

4.3.1 End of Life vs. Retrofit Measures

Site B. A customer replaced a cracked steam boiler with new direct fired gas heaters. The customer had used a leased portable boiler for a period until the new equipment was installed. The application used the old boiler efficiency as the baseline in a retrofit application.

The evaluator observed that in some cases, HVAC component replacement projects (boilers, non-boiler heaters, heating distribution, and roof replacement measures) were assigned a retrofit baseline where the evidence, as presented to the evaluator, indicated that the old equipment was at the end of its life. Table 4-5 summarizes and compares how the applicant and evaluator treated the base line.

Table 4-6: Measure Classification by Measure Type[3]

Installed component	Number of Sites	Applicant		Evaluator	
		Lost Opp	Retro	Lost Opp	Retro
Boilers	14	5	9	10	4
Direct fired units	3	1	2	3	
HVAC terminal distribution	3		3	1	2
Roof replacement	1	1		1	

Replacement of one or more boilers occurred at fourteen of the evaluated sites. The evaluator reclassified five sites that had been treated as retrofits by the applicant to a lost opportunity baseline. The decision to change a baseline from retrofit to lost opportunity was based on the customer's description of the state of the existing equipment and other compelling business factors. For example, the citation by the customer contact at Site C that there had been a business decision to change a plant from high pressure to low pressure steam in order to reduce mandatory staffing was a major factor in reclassifying this site as a lost opportunity site. A reclassification from a retrofit to lost opportunity base line did not always results in a drop in savings; Site D's base line was changed from 82% to 80% as a result of reclassification, which increased the savings.

The cases where the evaluator approved the retrofit equipment as the baseline for boilers were as follows:

- Water treatment plant where new boilers were installed but the existing boilers were retained as back-ups,
- Two public housing performance contracting project with multiple funding sources for energy purposes,

[3] The applicant tabulation is based on the baseline used in savings calculations and not on how the measure was classified in tracking. Tracking classification of Lost Opportunity vs. Retrofit was not always consistent with the baselines used to calculate savings.

Direct fired units were installed at three sites, one of which was a re-opening of an abandoned building, a second where heating was provided by a temporary boiler, and a third in a newly constructed building. The evaluator considered all three sites as lost opportunity projects.

The terminal unit replacements involved equipment that was not at the end of its useful life. In two cases, the evaluator agreed that the retrofit equipment was the baseline (a steam radiator relocation from monitor windows closer to occupants; a conversion from constant volume to a VAV system at a building constructed in 2002). The third site involved the installation of heat recovery on two existing air-handling units (treated by the evaluator as retrofit) and heat recovery on a new air-handler (treated by the evaluator as a lost opportunity).

The one roof replacement with added insulation in the project used code as the baseline with which the evaluator concurred.

4.3.2 Boiler burner replacements and controls

Site E. The vendor estimated the saving for the installation of boiler controls as 13.9% of the pre-installation gas use. Since the baseline measured combustion efficiency was 86%, the controls would have had to provide a 98% seasonal efficiency to achieve that level of savings.

Burner controls were installed at eight sites and achieve savings by sensing the oxygen levels in the combustion exhaust and precisely trimming the fuel-air mix to maintain optimum combustion. Four of the sites included conversion of oil-fired or dual-fired burners to gas fired burners. The combustion control vendor estimated usage savings ranged between 5 and 16% while the evaluated reductions were in the 1.5% to 10% range.

The evaluation baseline for the boiler control measures was one of the following:

- An efficiency vs. firing rate curve derived from multiple spot measurements of the boiler prior to installation of the controls. This is the preferred baseline when fuel switching has not occurred.
- An efficiency vs. firing rate curve from one spot measurement of the boiler prior to installation of the controls and an empirically derived curve fitted to that spot measurement;
- A code compliant combustion efficiency, using the empirically derived curve calibrated to the code efficiency. This is the baseline employed when fuel switching has occurred and reflects the fuel-switching policy.

The empirically derived curve was based on combustion measurements at multiple points in the firing range of 19 boilers at 12 different sites that were extracted from the evaluator's portfolio of linkage controlled boilers. The results show almost no variation in efficiency across the firing range.

4.3.3 High temperature output make-up air unit for space heating applications

Site B. A customer replaced steam-supplied unit heaters with no other mechanical exhaust in the building with 100% outdoor air make-up type units. The vendor claimed the unit saved energy, not only because of the high combustion efficiency, but also because the outdoor air through put for the unit was less than it would be for a 100% make-up air unit with a lower supply air temperature.

Three of the forty-eight sites had a particular vendor's direct fired heaters installed as high efficiency units in warehouse and light manufacturing facilities. These direct fired units had characteristics of:

- 92-100% combustion efficiency depending upon whether the unit was mounted inside or outside the heated space,
- 100% outdoor air,
- Relatively high supply air temperature of 160ºF (vs. 115 ºF)

While the combustion efficiency exceeds the code requirement (80%), these units introduce 100% outdoor air in a setting that is not typically ventilated at that rate. Since the unit operates on 100% outdoor air, there is the potential to use more energy than might be used by a vented or unvented code compliant heater. The manufacturer of the heater particularly emphasize that the high supply temperature permits a reduction of outdoor air, savings energy; however, the outdoor air volume should be driven by the needs of the space, not the capabilities of a particular system. Typically, a 100% outdoor air unit is used as a make-up air unit to balance an exhaust system.

In addition, the energy code requires units over a certain volume (currently, 5000 CFM with 70% or larger fresh air component). Two of the units observed during the evaluation met this criterion and were not equipped with heat recovery.

There was considerable discussion about this technology with the PA implementation and evaluation teams. There is agreement that an application for this technology should place a burden of proof upon the applicant to demonstrate a need for the level of ventilation provided by these units. It is not the job of the PA to ascertain code compliance for ventilation rates, however, the PA's also do not want to assert a technology is efficient because of one characteristic, such as combustion efficiency, but have it increase overall energy use due to another characteristic, such as the high volume of outdoor air.

There was considerable discussion concerning the interaction between infiltration and forced ventilation rates in very leaky building. It was proposed that at Site 229, the existing infiltration rates were so high (windows with missing panes in an old mill building), that there may not be a net increase in outdoor air.

4.3.4 Regenerative thermal oxidizing technology

Site F.The customer installed a new production line to meet expanded sales and selected regenerative technology for exhaust heat recovery. The equipment vendor estimated 148,000 therms of savings. The evaluator determined a realization rate of 103% for the site using a 70% efficient recuperative thermal oxidizer as the baseline.

Operations that release volatile organic compounds (VOCs) are required to capture the exhausted VOCs and incinerate them at temperatures exceeding 1,500ºF to render them environmentally harmless. The incineration options available to VOCs emitters are:

- No heat recovery from incineration. This option is not considered reasonable by the evaluator because the economics of at least some heat recovery are extremely compelling at the volumes observed for these sites.
- Recuperative thermal oxidizer with a stainless heat exchanger sized to provide about 70% efficiency. At this efficiency point, the recuperative technology is lower in cost than the regenerative. While higher recovery rates are possible with the recuperative technology; the costs increase beyond that of regenerative technologies at the production rates of these sites. Recuperative technology is more cost-effective in the mid-production range (15,000 to 30,000 CFM) at lower efficiency points.
- Regenerative thermal oxidizers use a ceramic medium that is 95% efficient, although the base cost of the unit is higher. As volumes increase, the regenerative become more cost effective. At about 30,000 CFM of continuous production, the regenerative is typically the most cost effective solution.

The evaluator conducted surveys of five manufacturers and distributors to determine an appropriate baseline for the RTO installation. The baseline selected for these sites represent a reasonable lowest cost option available to the customer, although not necessarily standard practice. Table 4-7 represents the findings of this market snapshot.

Table 4-7: Summary of RTO Vendor Survey

Low Flow <10,000 CFM Recuperative dominates market 50-65% efficient Lower capital costs Baseline: recuperative at 70%	Medium 10,000-30,000 CFM Both recuperative and regenerative share market At 70% recuperative efficiency, the two technology capital costs are equivalent Baseline: recuperative at 70%	Large >30,000 CFM Regenerative dominates market May not be the best choice for intermittent processes Baseline: regenerativ e

4.3.5 Industrial Process Measures

Six of the nine largest sites, all with savings over 100,000 therms, were at industrial sites. While potentially large savers, these measures are subject to the technical challenges and market swings of the underlying process. Some examples of projects illustrating these tensions follows:

> ***Site G**, a manufacturer of sheet metal housing components. The customer made a business decision to move an outsourced painting process in-house. The line, which included an incentivized infrared curing oven, operated in January 2010, but failed to meet certain temperature standards, so was shut down within a few weeks, The owner also ran into air emissions permitting problems. As of March 2012, the line was still not operating, although the customer was actively engaged in bringing the line up.*
>
> The evaluator calculated savings for this site (with downward adjustments) even though the line was not operating because the evaluator judged that the line would operate shortly and into the foreseeable future. The line was important to the customer's business strategy and the customer was confident the technical issues would be resolved.

__Site H__, a sewage treatment plant installed a suite of measures including new boilers capable of running on digester gas, boiler and building controls, and other measures. At the time of the site visit, the new boilers were off-line for a major system upgrade to the feedwater system.

The evaluation proceeded using other data on-site from which the boiler operation could be inferred and no savings penalty imposed, since the units were only temporarily off-line.

__Site I,__ a food processor expanded a line from a batch to continuous process. The continuous process permitted a change in temperatures based on the more sanitary closed continuous process. However, a new business partner required temperatures to be returned to the previous setting, eliminating much of the savings. Interestingly, plant production had doubled by the time of the evaluation and two more production lines were installed or in the process of being installed.

The evaluator utilized the higher production rates, however, savings were lower due to the temperature re-sets. The customer had no near term intentions to reintroduce the more efficient temperatures.

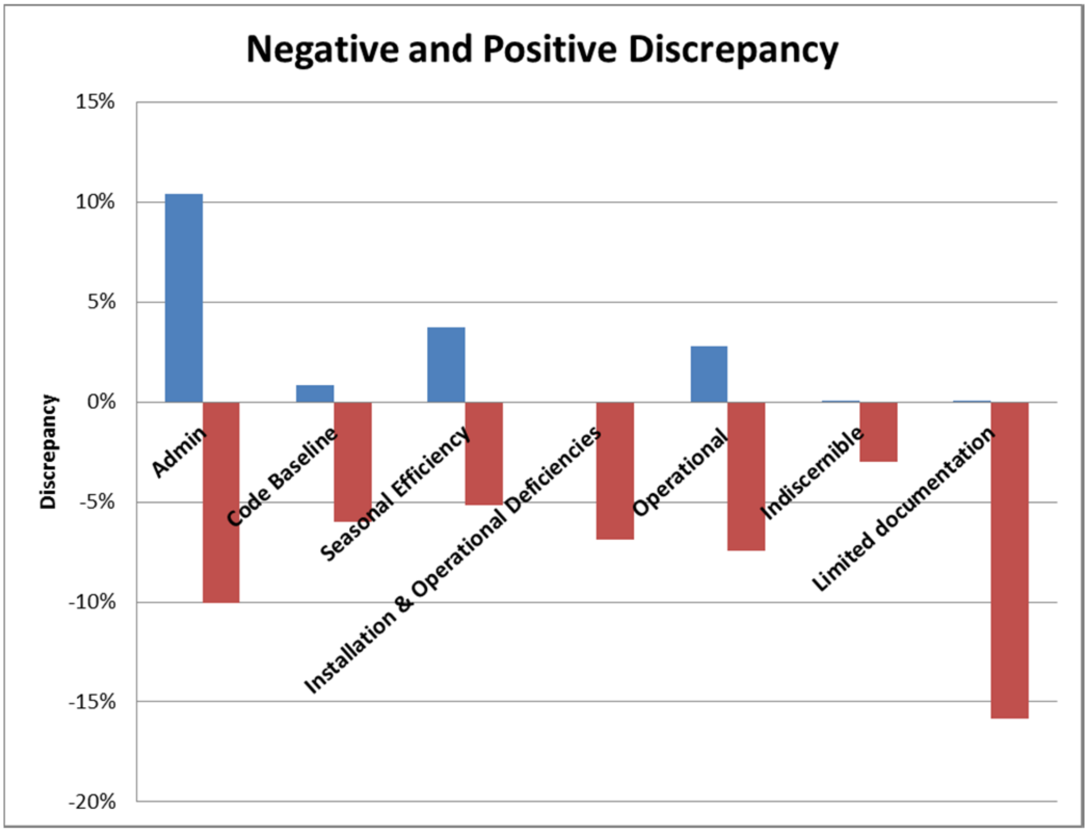
__Site J__, a manufacturing operation, added controls to a large central steam plant. After about two years of operation, the product was shipped overseas and the steam plant was shut down. In response, the PA changed the life of the measure to two years. The evaluator used customer records and billing data to conduct the analysis and computed typical first year savings, with the assumption that the PA's measure life change accounted for the plant shutdown.

These projects illustrate how challenging it can be to predict how long a process improvement may take to bring to full operation and how the business cycle may impact its use.

4.4 Discrepancy Analysis

Each site was reviewed to identify the factors that created discrepancies between evaluated and tracked savings and then to quantify and categorize them. While quantitative, this is not a statistically rigorous estimate of the factor on total program impact. The intent of this analysis is to provide the PAs with indicators of where they may want to focus future process improvements.

Figure 4-2: Weighted Contribution of Discrepancy



The graph shows both positive and negative impact of each discrepancy category. In brief, the categories are:

- Administrative –the savings reported in tracking do not match the paperwork. The Site K paperwork appeared to have the correct eQuest models and analysis, but the savings for the measure appeared to have been entered with an extra zero, increasing the savings by an order of magnitude in tracking.
- Code baseline –the applicant used the pre-installation equipment as the baseline, while the evaluator determined the equipment was at the end of life. Site D applicant specified a pre-installed efficiency of 82% (the old oil-fired efficiency), while the evaluator determined the boiler had been at the end of its useful life and specified a code baseline of 80% efficiency.
- Seasonal efficiency – this category captures the differences between the applicant and evaluator measured seasonal efficiencies. Frequently, particularly for condensing boilers, the applicant will use the peak efficiency (typically 94%), while the seasonal efficiency is typically 89-90%.

- Installation and operational deficiencies – captures where the installed equipment or operation as installed was deficient. At Site L, the customer had proposed heat recovery on a 450hp air compressor, but the installed unit was only 200hp, reducing the amount of heat available for space heating.
- Operational – accounts for discrepancies due to differences in applicant estimated and the evaluated hours of operation, full load equivalent hours, and the like. At Site M, the VAV system reduced the supply air lower than had been modeled resulting in additional savings.
- Limited documentation represents where the reasons for discrepancies cannot be attributed to a particular cause due to the brevity of the applicant's methodology. The applicant for Site N estimated the savings as 78% of the pre-installation usage for with an insulation measure, while the evaluator determined 55% savings. The reason for the difference between the two estimates cannot be determined.
- Indiscernible – captures what is essentially residual error between the applicant and evaluators estimates.

Each discrepancy was estimated independently as the difference in therms between the evaluated savings and what the savings would have been using the correct value. Residual error was categorized as "Indiscernible". The site's independent discrepancy values were reconciled to the site's evaluated discrepancy using the ratio of the sum of the individual results divided by the site's total evaluated discrepancy. An alternate method of calculating discrepancies is to cascade the calculations so that each subsequent calculation depends upon the results of the previous calculations. However, this method's results are highly dependent upon the ordering of the analysis, increasing the apparent impact of the first category, decreasing the subsequent category impacts. The cascading method makes it harder to discern the value of a proposed process improvement action.

Site weights were applied to individual site discrepancy results to provide an estimate of the value of the discrepancy in the population.

Table 4-8 summarizes the weighted effect of each category of discrepancy. Note that the outcome of the discrepancy characterization itself has no impact on realization rates but only explains the sources of the discrepancies.

Table 4-8: Site Realization Rates and Discrepancies

Site	Tracking Savings-Therms	Weight	RR%	Admin	Code Baseline	Seasonal Efficiency	Installation & Operational Deficiencies	Operational	indicernible	Limited Back-up Documentation
Net	**2,531,346**	**272**	**64%**	**0%**	**-5%**	**-1%**	**-7%**	**-5%**	**-3%**	**-16%**
011	1,000	7	35.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-64.4%
018	75,188	1	38.9%	0.0%	0.0%	0.0%	0.0%	-43.3%	0.0%	-17.8%
065	1,936	21	31.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-68.8%
095	3,900	21	33.8%	-87.3%	0.0%	21.1%	0.0%	0.0%	0.0%	0.0%
101	4,656	21	200.8%	129.2%	0.0%	0.0%	0.0%	0.0%	0.0%	-28.4%
109	6,082	7	56.9%	0.0%	0.0%	-43.2%	0.0%	0.0%	-0.1%	0.0%
121	15,440	7	52.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-48.0%
122	17,157	7	34.8%	0.0%	0.0%	-60.5%	0.0%	0.0%	-4.7%	0.0%
125	32,079	2	51.7%	0.0%	0.0%	0.0%	-48.0%	0.0%	0.0%	0.0%
128	37,544	2	68.9%	0.0%	0.0%	-24.6%	0.0%	-6.8%	0.3%	0.0%
130	63,538	2	71.3%	0.0%	-24.0%	-5.0%	0.0%	0.0%	0.0%	0.0%
132	148,187	1	107.4%	0.0%	0.0%	-2.1%	0.0%	19.6%	-10.1%	0.0%
133	161,400	1	130.9%	0.0%	0.0%	0.0%	0.0%	28.2%	2.7%	0.0%
134	259,820	1	79.3%	0.0%	0.0%	-4.0%	-17.0%	0.0%	0.0%	0.0%
135	288,449	1	81.8%	0.0%	0.0%	-6.3%	0.0%	-12.0%	0.0%	0.0%
136	23,400	1	55.1%	0.0%	0.0%	-5.0%	-30.0%	-10.0%	0.0%	0.0%
185	1,558	18	20.0%	0.0%	-18.0%	-7.0%	-55.0%	0.0%	0.0%	0.0%
207	3,965	18	98.7%	0.0%	0.0%	-1.0%	0.0%	0.0%	-0.3%	0.0%
219	5,990	18	61.9%	0.0%	0.0%	0.0%	0.0%	-10.0%	0.0%	-28.1%
229	9,095	18	22.4%	0.0%	-51.0%	25.7%	0.0%	-13.9%	-38.4%	0.0%
235	10,703	5	108.2%	0.0%	0.0%	8.0%	0.0%	0.0%	0.0%	0.0%
238	13,462	5	13.1%	0.0%	0.0%	0.0%	-60.6%	-24.7%	0.0%	0.0%
242	17,244	5	15.9%	-78.2%	0.0%	0.0%	0.0%	0.0%	-5.8%	0.0%
250	22,646	5	21.8%	-78.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
253	26,128	3	103.2%	-0.6%	0.0%	0.0%	0.0%	0.0%	3.8%	0.0%
258	39,277	3	67.4%	0.0%	0.0%	0.0%	-29.6%	-3.0%	0.0%	0.0%
257	37,912	3	23.2%	-3.5%	-111.2%	3.2%	0.0%	32.9%	1.8%	0.0%
259	41,180	3	33.4%	0.0%	0.0%	0.0%	-18.9%	-48.0%	0.0%	0.0%
264	59,720	2	157.4%	0.0%	0.0%	0.0%	0.0%	57.4%	0.0%	0.0%
267	72,395	2	44.6%	0.0%	-24.0%	-7.5%	-36.0%	25.0%	-13.0%	0.0%
268	131,953	2	111.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	11.4%
269	142,859	1	85.0%	0.0%	0.0%	-9.4%	0.0%	-5.5%	0.0%	0.0%
270	152,674	1	72.3%	-12.4%	0.0%	-6.8%	0.0%	0.0%	-8.5%	0.0%
275	1,028	9	114.3%	0.0%	0.0%	0.0%	-8.2%	22.5%	0.0%	0.0%
282	2,067	9	47.5%	-6.0%	0.0%	0.0%	0.0%	-8.0%	0.0%	-38.9%
286	2,889	9	71.6%	-2.0%	0.0%	0.0%	0.0%	-12.3%	0.0%	-14.2%
306	10,299	9	59.8%	8.8%	22.0%	-39.1%	0.0%	-31.9%	0.0%	0.0%
313	18,296	3	33.5%	0.0%	8.0%	0.0%	0.0%	-74.0%	0.0%	0.0%
315	23,969	3	29.8%	0.0%	0.0%	-21.1%	0.0%	-49.1%	0.0%	0.0%
317	30,000	3	18.5%	0.0%	0.0%	0.0%	0.0%	-15.0%	0.0%	-66.5%
318	31,134	3	34.8%	0.0%	0.0%	0.0%	-65.2%	0.0%	0.0%	0.0%
321	40,513	2	56.9%	1.8%	0.0%	0.0%	0.0%	25.0%	0.0%	-69.9%
323	42,239	2	22.6%	0.0%	0.0%	17.6%	-10.1%	-75.8%	-9.0%	0.0%
325	68,544	2	74.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-25.7%
326	110,892	1	53.5%	0.0%	-16.5%	-10.0%	0.0%	-19.0%	0.0%	0.0%
327	134,992	1	42.8%	0.0%	0.0%	-22.3%	0.0%	-17.0%	0.0%	-17.9%
332	6,232	5	78.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-22.0%
335	79,715	1	12.9%	0.0%	0.0%	-6.7%	0.0%	-80.1%	0.0%	0.0%

4.5 Results of Small Sites Savings Project File Review

A random sample of 20 sites with savings of less than a 1000 therms was selected for project file reviews. The results of the review are summarized in Table 4-9**Error! Reference source not found.**.The individual site results can be viewed in Appendix A.

The measures were predominated by shell measures including roof and wall insulation, as well as a mix of air-sealing, windows and door measures.

One of the PAs is in the process of changing implementation contractors so project information was unavailable for six of the sites.

Based on billing information included in the project files, two sites had savings of greater than 100% and only two sites of less than 25%.It is possible that the sites with savings greater than 100% had other gas accounts which would have addressed improbably high savings rates, however, these high savings should be viewed as red flags.

Table 4-9: Summary of Small Savings Site Project Review

PAs represented	
NSTAR	1
Columbia	7
NGRID	8
Berkshire	4

Measure type	
TRV	1
Wall or roof insulation	10
Other envelope	6
DHW or heating replacement	3

Savings fraction	
Number of sites included	8
Low	3%
High	309%
Average	76%

Information availability	
Records unavailable - contract issue	6
Readily reproduce savings	6

4.6 Evaluation Execution

This section reviews the execution of the evaluation and the factors that affected it, including timing and project data acquisition.

4.6.1 Timing

The 2010 impact evaluation was started early, with the kick-off meeting occurring in September 2011. The early start-up had multiple benefits:

- Extended metering. Many of the sites had loggers in place for four to eight weeks which afforded a larger temperature range for regression analysis and also increased the probability of capturing cold weather during the mild winter.

- Recruiting. The longer timeframe permitted a longer recruitment period for hard to reach customers and also for more back-up cycles when customers refused to participate.

- Capture cold weather. The winter was particularly mild, therefore having meters in place for many of the sites for the one cold spell in January was important. A later start may have missed the one period of sub-10ºF weather.

It had been hoped that the early start would lead to an early wrap-up of results; this did not happen. In retrospect, it may have been unrealistic to plan for earlier results given the necessity of metering during the heating season.

4.6.2 Gathering documentation

The project files and billing data can be critical elements in performing an evaluation and can also be time-consuming and burdensome for the PAs to gather. The project files can be voluminous. The PAs must sift through the documents, select those that appear to be relevant, and then scan and send the results to the evaluator. Electronic copies of native spreadsheets and building simulations are often left on project reviewer desktops and do not make it to retrievable locations.

The evaluators noted that the project files were often incomplete. While all of the project files included an application and offer letter, about 10% of the project files included little else. Only two of the eleven sites using building simulations as the basis of the analysis included the building simulation models. Of thirteen sites explicitly using excel spreadsheets, only six included the spreadsheets. The evaluators went to project developers to determine measure implementation details and savings algorithms for some of the larger projects, however these efforts typically did not yield any new information.

In two cases, a critical project document was only discovered after the site analysis had been completed and the PA went back to the project file after reading the report. Site A had important information included in another project file, which was only discovered after diligent searches by the PA after reviewing the final site report.   Consequently, the evaluators were able to reproduce the savings with confidence only about half of the time.

One indicator of the project file disarray was the absolute value of 20% due to administrative error contribution as shown in Figure 4-2, This error includes double counting of savings, incorrect building simulation model references, and other clerical errors. However, since the error was almost evenly split between negative and positive impacts, the net effect was low. One-line calculations, which were typically a deemed savings fraction applied to the annual gas usage, were used as the savings estimate for about a third of the sites. Sometimes this is the only practical means of calculating savings, for example for TRVs, however, for other measures a more rigorous approach should produce better savings estimates.

Billing data is often a critical factor in the results. It is particularly challenging to gather the correct billing data at locations with multiple accounts, or where there have been changes to billing account information. The results for Site A illustrated in Figure 1-1 looked much poorer before all of the billing was acquired. It was particularly challenging at this site because as part of the system upgrade, fifteen new gas meters were installed.

5. Conclusions and Recommendations

Overall, the Custom Gas program appears to be aggressively pursuing energy efficiency opportunities across a range of customers. This year's sample included a number of industrial process measures, unique heat recovery systems, and larger multi-measure projects. Below are major findings and recommendations that apply statewide.

5.1 Realization Rates

The study produced statewide results that are reliable (±9.0%) at 80% confidence. The precision levels were found to be: National Grid: ±17.4%, Columbia Gas: ±83.2%, and NSTAR: ±11.2%, which meet the criteria for applying individual PA results established in the Protocol memo of November 2010.

The results do not support application by measure, although on first glance, the statistical outcomes may indicate otherwise. The measure classification is currently not robust enough to support by measure realization rates. These results indicate performance trends, but are not considered reliable for planning.

5.2 Program Improvement Recommendations

The evaluation team reviewed project files, conducted detailed analysis of the information provided in the files, and quantified discrepancies analysis to make the recommendations of this section.

5.2.1 Capital equipment baselines

Major HVAC capital projects that are replacing essential components (boilers, heaters, or roofs) should usually be treated as end-of-life measures requiring code as baseline unless there are strong and documented reasons for using the existing equipment as the baseline. The existing equipment was used as the baseline by the evaluator in two types of cases in this evaluation:

- The existing boilers were retained as back-up in an industrial process while new dual fuel (natural and digester gas) boilers were installed at a sewage treatment plant. Since the old equipment was retained at full capacity (and was in use at the time of the evaluation), the baseline was determined to be the old equipment.

- Two public housing complexes had boilers replaced as part of a performance contract with multiple funding sources.

5.2.2 Boiler burner replacements and controls

The evaluator recommends that the PA require a pre-installation boiler efficiency measurement, since the savings partly depends upon how badly the pre-existing boiler is controlled. Sites projecting more than a 5% efficiency improvements or greater should be carefully reviewed by the PA technical reviewer.

Where fuel-switching occurs (with a burner replacement), the baseline is the MA building code as indicated by the fuel-switching policy. Future applications should be screened in this manner as well.

5.2.3 High temperature output make-up air unit for space heating applications

The evaluator recommends considering 100% outdoor air units only in those applications where it has been established that the air-flows are required to balance other exhaust loads or where there is another compelling reason to consider high outdoor air flow rates. The implementers should also be aware that the energy code requires heat recovery on any unit exceeding 5000 CFM with a 70% or more outdoor air component. While it is not the PA's responsibility to determine the outdoor air requirements for each application, it is recommended that the PA put the burden of proof upon the applicant to demonstrate the need for fresh-air before approving the applications.

5.2.4 Regenerative thermal oxidizing technology

The evaluator conducted surveys of five manufacturers and distributors to determine an appropriate baseline for this project and one other RTO installation. The baseline selected for these sites represent a reasonable lowest cost option available to the customer.

5.2.5 Industrial Process Measures

Industrial process measures present opportunity and challenges. Process measures often generate large savings, but also present technical and programmatic challenges.

- Use least cost reasonable baselines. Since process measures are not common and there are no governing building codes, the PA's must first establish baselines. The evaluators selected the least expensive option select at least some of the time, s the baseline and recommend this framework going forward. This approach somewhat maximizes the savings and leaves the question about 'what the customer would have done in the absence of the program' in the domain of program attribution.

- Include a separate commissioning task. The PAs should consider a commissioning phase for unique process measures. This permits the customer to get a large incentive up front, but defers the final payment and ready-for-evaluation status until the unit is fully operational. This approach provides some leverage with the customer and protection for the PA.

5.2.6 Savings Estimation Procedures and Initial Screening

The following recommendations are from the last evaluation, but still apply based on observations of this evaluation's sample selection. This is not surprising, given these sites were installed before the findings of the last evaluation.

- **Calibrate models to weather-normalized billed usage.** Tracking calculation methodologies ranged from building simulations to single line calculations. Performance contractor proprietary software was also used for tracking estimates in a number of cases. Bin analyses, single line calculations, and proprietary software should usually be calibrated to weather-normalized billing usage. The use of TMY3 weather data as the standard in the calculations provides the most representative weather data for annualizing savings and should be used for all weather-sensitive savings calculations. This might impact savings by 5-10%.

- **Use current billed usage to "sanity check" savings estimates.** A simple screening to examine the measure savings as a percent of billed usage can help identify incorrect billing usage and applicant analysis that may require further scrutiny. Benchmarks should be assigned each measure type.

- **Consider interactivity of all measures for project savings.** It is important that the interactivity between all measures in a project be considered in the TA study.

- **Include complete billing records in the files.** The project file should include a copy of the actual bill for the site (with the meter number) so that the measure location can be accurately identified. This is particularly important for multi-family housing complexes with multiple meters. The record should include a 12-18 month history snapshot at the time of the application and also at the time of the incentive payment.

- **Ensure TA studies and supporting calculations are stored for future evaluations.** The evaluation team was not provided with the TA savings spreadsheets or building simulations used to estimate the tracking savings for some projects. When the tracking savings calculations were not available, evaluators were unable to clearly identify where the source of the differences in energy savings estimates.

- **Cross-check TA report with corresponding analysis files.** Analysis files (building simulation or spreadsheets) should be provided by the TA engineer with the final TA report, and the PAs should make sure that the provided analysis outputs match the report text and screening tool. There were a few instances across PAs in which the savings values in the TA reports and analysis files did not match, causing difficulties in identifying how the tracking savings were developed. Tracking values need to be updated with each subsequent reanalysis of the project.

- **Consider commissioning procedures for control measures.** The PAs should consider instituting a Minimum Requirements Document (MRD) procedure that can be used by inspectors to verify that complex control measures, such as an EMS or heat recovery, are properly operating.

5.3 Evaluation Recommendations

The following recommendations concern future custom gas impact evaluations.

5.3.1 Evaluation execution

These recommendations concern procedures and planning related to a custom gas impact evaluation:

Timing. The evaluation should begin as early as possible to permit sufficient time to gather documentation and recruit customers for the November/December swing season metering and also multi-week metering for a large portion of sites.

Gathering project files. The PAs may wish to consider having the evaluation engineers review project files at the PA's offices and be responsible for the selection and scanning of project documents. The evaluators may be the best judge of the value of certain documents and it may relieve some of the PA administrative burden.

Billing data. The billing data is vital. The evaluator needs to take particular care to record meter numbers at the site and to walk the building perimeter to identify potential multiple accounts. The PAs also need to be prepared to search CIS records for multiple accounts at the same location and also to provide billing as needed.

5.3.2 2011 program year impact evaluation

The Massachusetts gas programs have undergone significant changes, including much larger program goals with new procedures to ensure improved applicant file reviews. The 2009 and 2010 impact evaluations produced similar gross realization rates of about 70% (without the outlier site in 2009). The electric PAs have attained higher gross realization rates for electric program components and it is likely that over time, the realization rates will rise for the gas program as well.

The PAs and EEAC consultants may wish to consider a new round of custom gas impact evaluations of the 2011 program year because it is an opportunity to:

- Reap the benefits of process improvements with an increased realization rate,
- Identify additional technology and policy issues (such as fuel-switching and process measure baselines) that can inform implementers about better estimation and baseline practices.

However, if the PA's do not believe that process improvements are sufficiently rooted in the 2011 program year to yield different results, a third consecutive impact evaluation may not be warranted.

Instead, the PAs and EEAC consultant may wish to conduct a 'borehole' test in the fall of 2012. The evaluator would proceed with a full sample design of the 2011 program, gather the program files, gather available billing data, and formulate first round M&V plans. In late-November, the evaluator would report on the status of the project files, applicant baselines, and an initial billing analysis. Using this information, the informed Evaluation Group can decide whether to proceed with the on-site component of an impact evaluation or to post-pone a full impact evaluation for at least another year. If the full impact does not proceed, it is likely that there will be valuable lessons learned for both evaluation and implementers. This detailed desk review can provide the basis for:

- Review of baseline best practices, particularly regarding end of life measures, fuel-switching, and plant reconfiguration.
- Baseline research for any new industrial measures.

5.3.3 Standard boiler field performance

The evaluators have had an opportunity to measure dozens of high efficiency and standard boilers with combustion controls in the field and have well characterized their operation. In almost all cases, the measured boilers are modulating. What is less well characterized is the field operation of a standard non-modulating, single or multi-staged firing boiler. It is possible that there are additional secondary losses with a staged rather than modulating boiler that could be measured and used to better characterize a standard code compliant boiler.

The evaluators recommend that the PAs and EEAC consultants consider a study to measure ten to twenty single and multi-staged code compliant boilers. The intent of the study would be to characterize the efficiency vs. firing rates, stand-by losses, cycling rates as a function of temperature, and purge losses associated with the cycling. Since the PAs do not incentivize standard boilers, there is not a natural population from which to sample. However, the evaluators believe sites could be recruited for measurement through channel partners.

Appendix A: Small Savings Sites Tabulated Results

Appendix B: Site Reports

KEMA

ers

Appendix A: Results of Small Site Savings Project File Reviews

Site and Measure ID				Savings Fraction		Type	Timing		Analysis			
SITE	PA	Measure	Tracking	Baseline bill	Savings Fraction	Retro or Lost Opps	Year of App approval	Year of Analysis	Building Sim:eQuest, Other	Spreadsheets: Table, Bin, Hourly	One line calcs	Reproduce Savings - No Kind-of Yes
S72	NSTAR	STEAM VALVES	793	257	309%	Retro	2009	2009	Specification on TRV	No	N	No
S12	COLUMBIA	Ind. Water Heater	610	2318	26%	Lost Opp		2009	NA	No	One line calcs	No
S14	COLUMBIA	Wall Insulation	525	1310	40%	Retro	2009	2008	eQUEST	No	One line calcs	Yes
S15	COLUMBIA	Ind. Water Heater	871	3312	26%	Lost Opp		2009	eQUEST	No	N	No
S16	COLUMBIA	Overhead Door	54	164	33%	Retro	2010	2010	eQUEST	No	N	No
S17	COLUMBIA	Early Heating Replac	337	2,755	12%	Retro	2010	2010	eQUEST	No	N	No
S18	COLUMBIA	Roof Insulation	442	NA		Retro	2010	2010	NA	No	One line calcs (Saves 26%	Yes
S19	COLUMBIA	Roof Insulation	160	99.5	161%	Retro	2010	2010	NA	No	N	No
S38	NGRID	ROOF INSULATION	770			Retro						No
S39	NGRID	WALL INSULATION	770			Retro						No
S40	NGRID	PIPE INSULATION	171			Retro						No
S41	NGRID	ROOF INSULATION	458			Retro						No
S42	NGRID	ROOF INSULATION	458			Retro						No
S43	NGRID	ROOF INSULATION	458			Retro						No
S44	NGRID	INSL_WALL	795	NA		Lost Opp	2010	2010	TRACE	and Table Spreadshe	N	Yes
S45	NGRID	DOOR_INSL_OVRHD	235	8382	3%	Retro	2010	2010	NA	Table Spreadsheet	N	Yes
S01	BERKSHIRE	Destratification Fans	132	NA		Retro	2010		NA	Table Spreadsheet	N	No
S03	BERKSHIRE	Airsealing	65	NA		Retro	2010		NA	Table Spreadsheet	N	yes
S04	BERKSHIRE	Storm windows	152	NA		Retro	2010		NA	Table Spreadsheet	N	yes
S07	BERKSHIRE	Insulation (Wall & Attic)	159	NA		Retro	2010	2010	Building performance test	No	N	No