Impact Evaluation of 2009 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, August 30, 2011

Table of Contents

List of Exhibits:

1. Executive Summary

This document summarizes the work performed by the KEMA team, lead by ERS and including contributions from Itron, 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 2009.

This is the first state-wide evaluation that has ever been conducted for 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 2012 mid-term modification planning, planning for the second three-year plan, 2011 reporting, and subsequent year reporting, unless replaced by results from a subsequent study. Particular attention was paid to how the tracking estimates were determined in order to provide the PAs with specific, actionable recommendations for future improvements of those estimates. 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 2009 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. Measures which were no longer offered by the program, such as combined heat and power, were excluded from the study.

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-four sites.

1.1.1 Sampling Strategy

Since this was the first time that realization rates were estimated for the Massachusetts gas programs, there was much to be learned. Design objectives included the ability to study realization rates for therms saved across a number of efficiency measure categories as well as by Program Administrator with a precision of 80% confidence and ±10% precision.

Tracking system data for projects completed in 2009 was collected from the PAs, consolidated, reviewed, and grouped into five major categories. In many cases, assigning projects to these major categories required making some assumptions based on the project title field in the tracking data sets.

Many potential sample designs were considered in an attempt to achieve the project goals. After discussions with the PAs and the EEAC consultant representative about evaluation objectives, it was agreed to proceed with a by PA, by measure category sample design with the intent of achieving precision targets by PA, but ensuring representation of all measure categories. The sampling unit was an individual measure. It was agreed on a sample size of forty-seven measures and assumed an error ratio of 0.8. This design was expected to produce good precision results state wide.

Prior to the commencement of monitoring and verification activities, a protocol was developed for applying realization rate results to tracking savings for annual reporting purposes (the "Protocol"). The intent of the Protocol was to establish the decision rules in advance of the actual calculations, to ensure fair coverage for all PAs.

1.1.2 Measure Savings Estimates

Following the final sample selection of 2009 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. Detailed measurement and evaluation plans were produced for each site and reviewed by PA study managers. In cases where the measure had been incorrectly classified or the customer refused to participate, the primary selection was replaced with a back-up site.

Each site was visited by an 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. Loggers were left in place for a minimum of two weeks. When possible, trend data was secured from the building automation system.

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 a few cases used as the primary means of determining savings impacts.

1.2 Findings and Results

The site level results include the measure estimates of savings and a quantitative breakdown of the factors that caused the realization rates to deviate from 100%.

Three dropped sites could not be replaced, resulting in a final sample size of forty-four sites. Two of the final sample sites required special treatment. One site was excluded from the aggregated results as unevaluable due to unique circumstances where neither gas billing information nor baseline condition information were robust enough to identify savings with any certainty. The second site, an outlier, was retained in the analysis, but had its case weight adjusted due to the high realization rate and the magnitude of the savings. The final analysis incorporated site findings for forty-three sites.

In preparation for analyzing the evaluation results collected for the Custom Gas sample points, the original 2009 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 1-1.

Table 1-1: Summary of Custom Gas Results

Statistic	Annual Therms
All Program Administrators	
Total tracking savings	1,982,441
Total measured savings	1,736,323
Realization rate	87.8%
Relative precision at 80% confidence	±10.8%
Error bound at 80% confidence	186,955
Sample size	43
Error ratio	1.93

Table 1-2 presents results by measure category.

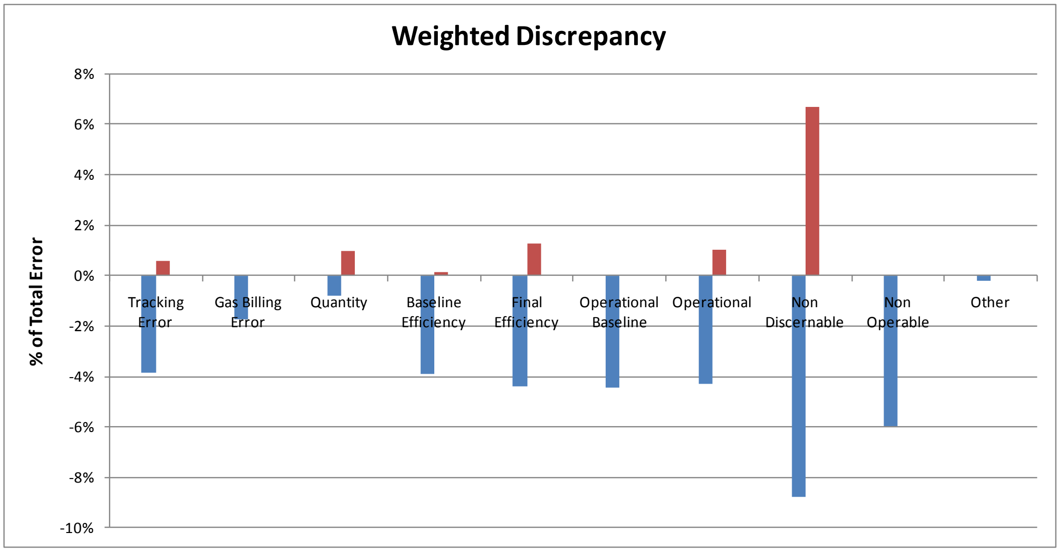
Table 1-2: Results by Measure Category

Measure	Annual Therms
Boilers	
Total tracking savings	408,746
Total measured savings	176,620
Realization rate	43.2%
Relative precision at 80% confidence	±29.4%
Error bound at 80% confidence	51,897
Sample size	9
Error ratio	0.84
Controls	
Total tracking savings	460,645
Total measured savings	570,889
Realization rate	123.9%
Relative precision at 80% confidence	±21.6%
Error bound at 80% confidence	123,430
Sample size	9
Error ratio	3.26
Insulation	
Total tracking savings	221,256
Total measured savings	113,240
Realization rate	51.2%
Relative precision at 80% confidence	±9.8%
Error bound at 80% confidence	11,057
Sample size	9
Error ratio	0. 27

Measure	Annual Therms
Non-boiler heating	
Total tracking savings	408,371
Total measured savings	383,469
Realization rate	93.9%
Relative precision at 80% confidence	±15.7%
Error bound at 80% confidence	60,112
Sample size	9
Error ratio	0.43
Other	
Total tracking savings	483,423
Total measured savings	539,002
Realization rate	111.5%
Relative precision at 80% confidence	±11.7%
Error bound at 80% confidence	63,119
Sample size	7
Error ratio	0.44

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. Figure 1-1 summarizes the discrepancy impacts, excluding the Outlier site.

Figure 1-1 Weighted Contribution of Discrepancy



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 (±10.8%). The precision levels for National Grid and Columbia Gas were ±13.2% and ±19.1% respectively, which are reasonably good and meet the criteria for applying individual PA results established in the Protocol.

The results do not support application by measure, although on first glance, the statistical outcomes may indicate otherwise. However, 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.

The PAs may also wish to consider whether to exclude the outlier site when planning. While the outlier impact is included in the retrospective analysis and realization rates presented in Table 1-1, the likelihood of another high gas usage site with grossly underestimated savings occuring in the future is low and a more accurate predictor of future savings may be represented by the statewide realization rate of 71.3% with the Outlier excluded. Likewise, the most appropriate error ratio for sample planning may be 0.68 which is the error ratio with the Outlier excluded.

1.3.2 Tracking Savings Estimation 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. These are grouped by three stages of application processing.

Initial Screening

- Major capital projects should typically be treated as equipment replacements with code as the baseline, since, typically, customers consider these measures only at or near the end of life of the equipment. Consider including investigation of early replacement in future impact studies.
- Institute measure classification protocols to ensure consistency for application of measure-level realization rates in the future.
- Require technical assistance study providers to calibrate models to weather-normalized billed usage to ensure the models represent typical savings for the building where the measure was installed.
- “Sanity check” savings estimates using the customer’s bills
- Consider interactivity of all measures for project savings.

PA Maintained Project Documentation

- Include complete billing records, including a bill with the meter number and all the available monthly billing history.
- Ensure TA studies, analytical files in native form, and supporting calculations are stored for future evaluations.

Post Implementation

- Consider requiring commissioning procedures for control measures.
- Ensure the final tracking estimate is correct with a cross-check of the final technical assistance report and site inspection results.

2. Description of Sampling Strategy

The primary focus of the sample design was to examine various precision scenarios for the Custom Gas programs in Massachusetts. Since this was the first time that realization rates were estimated for the Massachusetts gas programs, there was much to be learned. Design objectives included the ability to study realization rates for therms saved across a number of efficiency measure categories as well as by Program Administrator. However, given the lack of knowledge about the site-to-site variability of realization rates for gas measures and the small sample sizes allotted to this study, there was general agreement that the results by PA and/or measure category may not be statistically reliable, although reasonably good precision was expected statewide, over all measures.

Prior to the commencement of monitoring and verification activities, a plan or "Protocol" was developed for applying realization rate results to tracking savings for annual reporting purposes. The Protocol prescribes calculating precisions for each PA, and for each PA and measure combination in isolation of the remainder of the sample. The Protocol further states the results should be used for any segment meeting specified precision thresholds and minimum numbers of sample points. The plan is documented in a memorandum included in Appendix A . The intent of the Protocol was to establish the decision rules in advance of the actual calculations, to ensure fair coverage for all PAs.

Tracking system data for projects completed in 2009 was collected from the PAs, consolidated, reviewed, and grouped into five major categories:

1. Boilers
2. Controls
3. Insulation
4. Non-boiler heating
5. Other

In many cases, assigning projects to these major categories involved making some assumptions based on the project title field in the tracking data sets provided by the PAs. The resulting study population of 339 Custom Gas projects is summarized in Table 2-1. Seven measures had zero savings.

Table 2-1: Population Statistics

Measure Category	Projects	Total Therms	Average Therms	Minimum	Maximum	StdDev	CV
Boilers	71	408,746	5,757	175	56,950	8,946	1.55
Controls	29	525,232	18,111	73	154,050	31,689	1.75
Insulation	99	221,256	2,235	39	58,247	6,143	2.75
Non-boiler heating	83	408,371	4,920	0	71,988	11,409	2.32
Other	57	483,423	8,481	0	88,000	15,007	1.77
Total	339	2,047,028					

Information about the distribution of these projects across Sponsors is included as Table 2-2 and the therms savings associated with these projects are shown in Table 2-3. While Columbia Gas sponsored over half of the projects in 2009, National Grid's programs accounted for over half of the therms saved.

Table 2-2: Population Distribution by PA

Program Administrator	Boilers	Controls	Insulation	Non-Boiler Heating	Other	Total
National Grid	32	7	16	9	28	92
NSTAR	1	2	7	11	20	41
New England Gas	0	0	0	2	2	4
Berkshire Gas	0	6	8	0	0	14
Columbia Gas	38	12	68	55	6	179
Unitil	0	2	0	6	1	9
Total	71	29	99	83	57	339

Table 2-3: Savings Distribution by PA

Program Administrator	Boilers	Controls	Insulation	Non-Boiler Heating	Other	Total
National Grid	286,596	355,456	58,077	96,186	347,844	1,144,159
NSTAR	9,201	6,110	23,486	90,899	58,416	188,112
New England Gas	0	0	0	211	0	211
Berkshire Gas	0	1,342	8,768	0	0	10,110
Columbia Gas	112,949	162,093	130,925	219,412	75,363	700,742
Unitil	0	231	0	1,663	1,800	3,694
Total	408,746	525,232	221,256	408,371	483,423	2,047,028

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). Since there were no prior studies from which to estimate this relationship, the preliminary sample planning was done for error ratios ranging from 0.6 to 1.0.

Many potential sample designs were considered in an attempt to achieve the project goals of 80% confidence and ±10% precision by measure category. The overall sample size was limited to forty sites. An initial design was selected to achieve the best possible precision for the five measure categories, based on the population data provided by the Sponsors and an assumption that the error ratio for the realization rate will be 0.80. Table 2-4 shows the stratum cut points and distribution of sample sites in this design.

Table 2-4: Sample Design based on Five Measure Categories

Measure Category	Stratum	Maximum Therms	Projects	Total Therms Saved	Planned Sample	Inclusion Probabilities
Boilers	1	6,017	50	100,333	3	0.0600
Boilers	2	14,783	15	126,671	3	0.2000
Boilers	3	56,950	6	181,742	3	0.5000
Controls	1	14,677	21	84,276	2	0.0952
Controls	2	37,589	4	111,182	2	0.5000
Controls	3	73,305	3	175,724	2	0.6667
Controls	4	154,050	1	154,050	1	1.0000
Insulation	1	1,256	67	40,295	3	0.0448
Insulation	2	3,843	22	49,892	2	0.0909
Insulation	3	14,139	9	72,822	2	0.2222
Insulation	4	58,247	1	58,247	1	1.0000
Non-boiler heating	1	4,187	65	91,699	3	0.0462
Non-boiler heating	2	21,200	14	107,960	3	0.2143
Non-boiler heating	3	71,988	4	208,712	2	0.5000
Other	1	6,207	41	94,493	3	0.0732
Other	2	27,122	10	112,468	2	0.2000
Other	3	44,470	5	188,462	2	0.4000
Other	4	88,000	1	88,000	1	1.0000

Table 2-5 lists the calculated precision estimates for this design, following stratification. While none of the individual measure categories come close to achieving the desired relative precision of ±10%, the overall precision comes reasonably close at ±14.47%.

Table 2-5: Estimated Precision for Five Measure Categories

Measure Category	Projects	Total Therms Saved	Error Ratio	Confidence Level	Planned Sample Size	Anticipated Relative Precision
Boilers	71	408,746	0.8	80%	9	±32.31%
Controls	29	525,232	0.8	80%	7	±27.96%
Insulation	99	221,256	0.8	80%	8	±32.75%
Non-boiler heating	83	408,371	0.8	80%	8	±34.29%
Other	57	483,423	0.8	80%	8	±31.96%
Total	339	2,047,028	0.8	80%	40	±14.47%

2.2 Final Sample

Following this design, the allocation of sample points across measure categories and PAs was reviewed. The optimal allocation of the forty sample points, followed by rounding to whole numbers of sites, resulted in some cases in which PAs were not represented in measure categories where they had projects. In order to ensure adequate representation, additional sample points were added to each cell that contained projects but received no allocation. The final number of sample points that were drawn for each PA and measure category is shown in Table 2-6.

Table 2-6: Final Sample Selection by Sponsor

Program Administrator	Boilers	Controls	Insulation	Non-Boiler Heating	Other	Total
National Grid	5	4	1	1	4	15
NSTAR	1	1	1	2	1	6
New England Gas	0	0	0	1	1	2
Berkshire Gas	0	1	1	0	0	2
Columbia Gas	3	3	6	5	2	19
Unitil	0	1	0	1	1	3
Total	9	10	9	10	9	47

After discussions with the PAs and the EEAC consultant representative about evaluation objectives, KEMA recommended and the PAs and EEAC consultant representative agreed that they proceed with this sample design for forty-seven sites in Massachusetts. This design was expected to produce good precision results state wide and reasonable precision results for National Grid and Columbia Gas individually.

Table 2-7 presents the final sample disposition. There were several reasons some sites were dropped. These reasons are summarized in Table 2-8. Ultimately, five sites were dropped and replaced, while five sites were dropped with no replacement. The excluded site was determined to be unevaluable and is discussed below. The over-sampled site was thought to be a primary site, but was really a potential replacement site. Since site work was already initiated on this site, it was left in the final sample. A National Grid 'Other' site was selected as a replacement for an 'Other' measure after the Unitil and New England Gas dropped sites eliminated two 'Other' sites.

Table 2-7: Final Sample Disposition

Program Administrator	Original Sample	Dropped & Replaced	Dropped - No Alternative	Un-evaluable	Added	Final
Columbia Gas	19	4				19
Berkshire Gas	2					2
New England Gas	2		2			0
National Grid	15			1	1*	15
NSTAR	6	1			1	7
UNITIL	3		3			0
Total	47	5	5	1	2	43

*Selected as an 'Other' measure replacement.

Table 2-8: Reasons for Dropped Sites

Program Administrator	Mis-classified	Prescriptive Measure	Site Refused	Un-evaluable	Final
Columbia Gas	3		1		4
Berkshire Gas					0
New England Gas		2			2
National Grid				1	1
NSTAR			1		1
UNITIL		3			3
Total	3	5	2	1	11

Unevaluable Site. The measure intended for evaluation of Site 337 was the distribution loss savings due to a conversion of an oil-fired steam boiler and distribution system to a gas-fired hydronic system. The reduction in steam losses would be due to the elimination of steam traps and the related losses from steam trap failures. There was no documentation of the ex-ante system condition; therefore savings estimates would have been based on typical practice and would not reflect any specific information gathered on-site about those conditions. Billing analysis was not an option because the facility had expanded by 25% concurrent with the measure installation.

The engineering team performed an uncertainty analysis using propagation of error relationships to estimate the overall uncertainty associated with the evaluated savings. Accuracy bounds were assigned to each of eight variables in the energy savings equation and overall uncertainty was calculated. The resulting overall accuracy was estimated to be -63% to 97%, a range well beyond the normal bounds of normally acceptable evaluation savings estimates. Because the uncertainty was so great, the evaluation team recommended removing the site from the sample and excluding the results from the aggregate realization rate analysis.

Discussions were held with the EEAC Evaluation Consultant. Concerns were expressed by the Consultant about setting a precedent for eliminating sites with poor documentation, as this may create a positive bias in the results and indirectly reward PAs for poor processes. However, given the nature of the site and the fact that this effort was an initial evaluation round, it was agreed the site would be excluded. In the future, a lack of documentation will not provide sufficient reason to exclude a site or assign a realization rate of 100%. This does not exclude the possibility of using an alternate site if available and appropriate and if time allows.

Table 2-9 summarizes the final sites for which monitoring and verification activities were completed.

Table 2-9 Strata by Site ID for Final Selection

Site ID	PA	Stratum	Evaluator	Sample Measure Description	Verified Measure Description	Building Type
72	Berkshire Gas	1	ERS	Controls	Programmable thermostat, conventional space heating	Mixed use
179	Berkshire Gas	1	KEMA	Insulation	Added insulation as part of reroofing	Office
31	Columbia Gas	1	ERS	Boilers	Condensing boiler replacement, space heating	Apartment building
40	Columbia Gas	1	ERS	Boilers	Condensing boiler replacement, space heating	Elementary school
64	Columbia Gas	2	ERS	Boilers	Condensing boiler replacement, space heating	Elementary school
85	Columbia Gas	1	Itron	Controls	EMS boiler sequencing, conventional space heating	Mixed use, very large space
89	Columbia Gas	1	Itron	Controls	EMS HVAC unit controls, conventional space heating	Mixed use, very large space
93	Columbia Gas	2	ERS	Controls	EMS boiler sequencing and HVAC unit controls, conventional space heating	Elementary school
147	Columbia Gas	1	KEMA	Insulation	Membrane roof insulation, retrofit	Office
150	Columbia Gas	1	ERS	Insulation	Window replacement	Public housing
163	Columbia Gas	1	ERS	Insulation	Window replacement	Public housing
171	Columbia Gas	2	ERS	Insulation	Window replacement	Public housing
174	Columbia Gas	2	KEMA	Roof Insulation	Membrane roof insulation, retrofit	Office
199	Columbia Gas	4	ERS	Insulation	Roof insulation with fill, retrofit	Apartment building
202	Columbia Gas	1	ERS	Non-boiler heating	Indirect DHW component of a boiler replacement	Lodging - retreat center
235	Columbia Gas	1	ERS	Non-boiler heating	Indirect DHW component of a boiler replacement	Public housing
267	Columbia Gas	2	ERS	Non-boiler heating	Indirect DHW component of a boiler replacement	Public housing
270	Columbia Gas	2	ERS	Non-boiler heating	Indirect DHW component of a boiler replacement	Public housing
279	Columbia Gas	3	ERS	Non-boiler heating	Process - new dry-cleaning machines	Dry cleaners and laundry
324	Columbia Gas	2	ERS	Other	Occupancy controlled ventilation, process paint booth	Small industrial
332	Columbia Gas	2	ERS	Other	CO controlled ventilation, conventional heating	High school
44	National Grid	1	ERS	Boilers	Condensing boiler new construction, DHW	Public housing
61	National Grid	2	ERS	Boilers	Boiler O_2 trim controls, oil to gas burner upgrade	University dormitory

Site ID	PA	Stratum	Evaluator	Sample Measure Description	Verified Measure Description	Building Type
66	National Grid	3	ERS	Boilers	Condensing boiler replacement, space heating and DHW	Public housing
68	National Grid	3	ERS	Boilers	Boiler O_2 trim controls, low pressure steam for DHW	Lodging - large hotel
71	National Grid	3	ERS	Boilers	Boiler O_2 trim controls, steam boilers for food processing	Industrial food processing
96	National Grid	2	Itron	Controls	EMS HVAC unit controls, new construction conventional space heating	High school
98	National Grid	3	ERS	Controls	TRV, conventional space heating	Public housing
99	National Grid	3	Itron	Controls	EMS HVAC unit controls, process ventilation	Paint manufacturing
100	National Grid	4	ERS	Controls	TRV, conventional space heating	Public housing
195	National Grid	3	KEMA	Insulation	Window replacement	Nursing home
282	National Grid	3	ERS	Non-boiler heating	Reflective roof, insulation, HVAC redesign, new construction	Pharmaceutical manufacturer
300	National Grid	1	ERS	Other	Ventilation heat recovery, new construction	Residential treatment
305	National Grid	1	ERS	Other	Ventilation heat recovery, replacement	Office - gym
337	National Grid		Itron	Other	Steam to hydronic system conversion, conventional heating	High school
338	National Grid	3	ERS	Other	Supermarket refrigeration heat recovery	Grocery store
339	National Grid	4	ERS	Other	Steam venting reduction, food processing	Industrial food processing
62	NSTAR	2	ERS	Boilers	Condensing boiler replacement, space heating	Office building
81	NSTAR	1	ERS	Controls	EMS boiler sequencing and HVAC unit controls, conventional space heating	Elementary school
176	NSTAR	2	KEMA	Insulation	Wall insulation, retrofit	Warehouse
256	NSTAR	1	ERS	Non-boiler heating	Tankless DHW, laundry usage	Dry cleaners and laundry
266	NSTAR	2	ERS	Non-boiler heating	Condensing boilers, DHW	Apartment building
278	NSTAR	2	ERS	Non-boiler heating	Condensing boiler replacement, space heating and DHW	Apartment building
318	NSTAR	1	ERS	Other	TRV	Apartment building

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 2009 Custom Gas applications and prior to beginning any site visits, ERS developed detailed measurement and evaluation plans for each of the forty-four 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) valves, were evaluated using a billing analysis. Billing analysis was used if the baseline conditions could not be confirmed and no other significant changes had occurred at the site. In a billing analysis, pre-implementation and post-implementation billed gas use is analyzed using historical weather and TMY weather data files.

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 C.

3.3 Analysis Procedures

In order to aggregate the individual site results from the Custom Gas sample, KEMA applied the model-assisted stratified ratio estimation methodology.[12] 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.

The results presented in the following section include realization rates (and associated precision levels) for annual therms savings.

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.

---

[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.

4.1 Site Level Results and Analysis

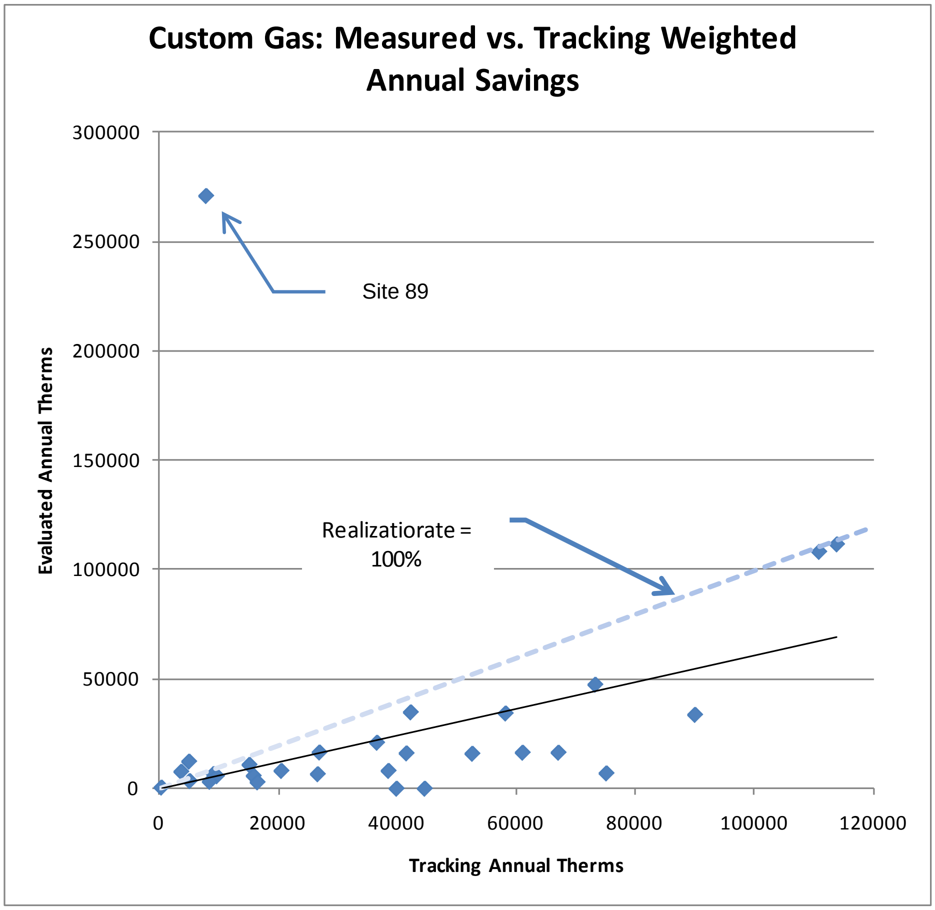
The key result of the study is the evaluated typical annual gas savings and the site realization rate, which are shown in Table 4-1. Additional quantitative analysis was conducted to determine why the evaluator's savings differed from that of the applicant's. A quantitative but not statistically rigorous estimate of that factor's impact on total program savings is noted in the last row of the table. Section 5 discusses the findings related to the discrepancy analysis.

Table 4-1 Site Realization Rates and Discrepancies

Site ID	PA	Tracking Savings-MMBtu	RR%	Tracking	Gas Billing	Quantity	Baseline Efficiency	Final Efficiency	Oper-ation Baseline	Oper-ation	Indiscern-able	Non Operable	Other
31	Columbia Gas	172	25%	0%			-52%	-9%			-15%		
40	Columbia Gas	340	30%	-2%	-37%			-11%			-20%		
44	National Grid	474	37%	-52%				-16%			5%		
61	National Grid	873	27%				-5%	-48%			-20%		
62	NSTAR	920	73%					29%			-56%		
64	Columbia Gas	959	25%				-24%	-16%			-35%		
66	National Grid	1929	21%	-41%				-20%			-18%		
68	National Grid	1997	0%									-100%	
71	National Grid	5695	98%				-6%	18%		-13%	-1%		
72	Berkshire Gas	7	119%								19%		
81	NSTAR	254	244%								144%		
85	Columbia Gas	558	0%									-100%	
89	Columbia Gas	794	3412%							3312%			
93	Columbia Gas	2080	39%		-27%			0%	-26%		-8%		
96	National Grid	3759	9%						-91%				
98	National Grid	4231	82%								-18%		
99	National Grid	7331	65%							-29%	-6%		
100	National Grid	11091	98%								-2%		
147	Columbia Gas	86	72%			-8%				-5%	-11%		-5%
150	Columbia Gas	90	37%			47%	6%			-78%	-39%		
163	Columbia Gas	117	39%			31%	6%			-68%	-30%		
171	Columbia Gas	142	37%			40%				-71%	-32%		
174	Columbia Gas	163	59%								-41%		
176	NSTAR	173	67%								-33%		
179	Berkshire Gas	207	18%				-77%				-5%		
195	National Grid	733	57%			-36%							-7%
199	Columbia Gas	5825	59%								-41%		
202	Columbia Gas	18	206%	284%							-178%		
235	Columbia Gas	126	61%							-5%	-33%		
256	NSTAR	274	10%	-25%			-23%	-22%		-11%	-10%		
266	NSTAR	494	23%					-12%		-45%	-21%		
267	Columbia Gas	517	94%								-6%		
270	Columbia Gas	517	94%								-6%		
278	NSTAR	2120	43%				-37%	-4%			-16%		
279	Columbia Gas	4100	141%								41%		
282	National Grid	7199	99%					-1%		0%	0%		
300	National Grid	179	274%								174%		
305	National Grid	222	0%									-100%	
318	NSTAR	401	118%								18%		
324	Columbia Gas	645	300%							48%	150%		
332	Columbia Gas	1688	31%									-69%	
338	National Grid	4447	101%					0%		1%	0%		
339	National Grid	8800	127%							12%	15%		
				-3%	-2%	0%	-4%	-3%	-4%	-3%	-2%	-6%	0%

Figure 4-1 presents a scatter plot of evaluation results for annual therm savings using all PA sample points. 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 pretty 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



4.1.1 Outlier Site

The measure evaluated at this site, Site 89, was an energy management system (EMS) upgrade. This measure replaced the existing front-end control software and computer of an older, inoperable system. The specific evaluated strategy scheduled approximately 120 air handling units (AHUs) for occupied and unoccupied operation, as opposed to running constantly in occupied mode. The facility is a very large mixed-use building of over 2 million square-feet currently running 75% vacant. The high vacancy rate drives energy savings because the AHUs are operated only as necessary to maintain minimum space temperatures. There were other challenges associated with evaluating this site, including the vacancy rate, but the focus of this discussion is on the use of evaluated results and not the method for estimating savings.

This site has a large impact on the outcome of the statewide, the particular PA, and the measure realization rates and precisions. This site had significant savings to begin with, but the high realization rate compounds the statistical impact of the site.

The evaluation team concluded that it was not appropriate to leave Site 89 results "as is" in the aggregated result because the site unduly influences the outcome. Removing the site drops the realization rate by almost 50%. The team also concluded that it was not appropriate to exclude the site because its savings made a contribution to the portfolio that should be counted. The team took a third path and re-classified the site as a census site, changing its case weight from 4.5 to 1 and adjusting the case weights of its previous strata. This approach captures the savings of the site, but does not propagate an unrealistic realization rate across the other sites in the strata. Table 4-2 summarizes the outlier disposition choices.

Table 4-2 Outlier Analysis

	Results for Outlier Options		
	Include	Exclude	Adjust Weight
Statewide Realization Rate	144.3%	71.3%	87.8%
% Precision	±55.0%	±11.1%	±10.8%
Columbia Gas Realization Rate	±268.5%	±62.4%	±108.9%
% Precision	83.2%	24.1%	19.1%
Control Measures Realization Rate	±352.8%	±56.8%	±123.9%
% Precision	87.7%	21.4%	21.6%

4.2 Retrospective Realization Rates

In preparation for analyzing the evaluation results collected for the Custom Gas sample points, the original 2009 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 4-3

Table 4-3: Custom Gas Case Weights

Measure	Program Administrator	Stratum	Total Projects	Total Annual Therms	Projects in Sample	Case Weight
Boilers	Columbia Gas	1	31	57,038	2	15.50
Boilers	Columbia Gas	2	7	55,911	1	7.00
Boilers	National Grid	1	19	43,295	1	19.00
Boilers	National Grid	2	7	61,559	1	7.00
Boilers	National Grid	3	6	181,742	3	2.00
Boilers	NSTAR	4	1	9,201	1	1.00
Controls	Columbia Gas	1	8	62,954	1	8.00
Controls	Columbia Gas	2	1	7,936	1	1.00
Controls	Columbia Gas	3	2	52,534	1	2.00
Controls	Columbia Gas	4	1	38,669	0	0.00
Controls	Berkshire Gas	1	6	1,342	1	6.00
Controls	National Grid	1	2	5,703	0	0.00
Controls	National Grid	2	2	58,648	1	2.00
Controls	National Grid	3	2	115,613	2	1.00
Controls	National Grid	4	1	110,905	1	1.00
Controls	NSTAR	1	2	6,110	1	2.00
Controls	Unitil	1	2	231	0	0.00
Insulation	Columbia Gas	1	53	33,408	3	17.67
Insulation	Columbia Gas	2	12	27,222	2	6.00
Insulation	Columbia Gas	3	2	12,048	0	0.00
Insulation	Columbia Gas	4	1	58,247	1	1.00
Insulation	Berkshire Gas	1	8	8,768	1	8.00
Insulation	National Grid	1	7	4,278	0	0.00
Insulation	National Grid	2	4	10,263	0	0.00
Insulation	National Grid	3	5	43,536	1	5.00
Insulation	NSTAR	1	2	759	0	0.00
Insulation	NSTAR	2	3	5,489	1	3.00
Insulation	NSTAR	3	2	17,238	0	0.00

Measure	Program Administrator	Stratum	Total Projects	Total Annual Therms	Projects in Sample	Case Weight
Non-boiler heating	Columbia Gas	1	43	62,769	2	21.50
Non-boiler heating	Columbia Gas	2	10	62,565	2	5.00
Non-boiler heating	Columbia Gas	3	2	94,078	1	2.00
Non-boiler heating	New England Gas	1	2	211	0	0.00
Non-boiler heating	National Grid	1	7	9,733	0	0.00
Non-boiler heating	National Grid	2	1	14,465	0	0.00
Non-boiler heating	National Grid	3	1	71,988	1	1.00
Non-boiler heating	NSTAR	1	7	17,323	1	7.00
Non-boiler heating	NSTAR	2	3	30,930	2	1.50
Non-boiler heating	NSTAR	3	1	42,646	0	0.00
Non-boiler heating	Unitil	1	6	1,663	0	0.00
Other	Columbia Gas	1	1	2,296	0	0.00
Other	Columbia Gas	2	4	40,988	2	2.00
Other	Columbia Gas	3	1	32,079	0	0.00
Other	New England Gas	1	2	0	0	0.00
Other	National Grid	1	17	31,981	2	8.50
Other	National Grid	2	6	71,480	0	0.00
Other	National Grid	3	4	156,383	1	4.00
Other	National Grid	4	1	88,000	1	1.00
Other	NSTAR	1	20	58,416	1	20.00
Other	Unitil	1	1	1,800	0	0.00

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 4-4 summarizes the state-wide results of this analysis. The realization rate for Custom Gas measures was found to be 87.8%. The relative precision for this estimate was found to be ±10.8% at the 80% level of confidence. The error ratio was found to be 1.93.

This incongruously good precision and poor error ratio is largely due to the Outlier. Changing the outlier's weight to one ("1") results in a big improvement in the relative precision because this site is self representing and no longer contributes to the overall variance of the estimate. However, the error ratio calculation is still negatively impacted by the fact that an observation like this sample point is possible and so very different from the others, in terms of the estimated relationship between evaluated and tracking savings. If the outlier site is excluded, which may be appropriate for future study planning, the error ratio was 0.68.

Table 4-4: Summary of Custom Gas Results

Statistic	Annual Therms
All Program Administrators	
Total tracking savings	1,982,441
Total measured savings	1,736,323
Realization rate	87.8%
Relative precision at 80% confidence	±10.8%
Error bound at 80% confidence	186,955
Sample size	43
Error ratio	1.93

The results of KEMA's analysis of realization rates by PA follow in Table 4-5. It was anticipated that only National Grid and Columbia Gas would have enough sample points to produce estimates with adequate precision to use their individual results, in accordance with the Protocol. National Grid's precision came in at ±13.2%, which is very close to the target of 10%. However, Columbia Gas's precision was higher, at ±19.1%, due to the influence of a large outlier. Table 4-6 compares each of these PA's rates to the remainder of the state. A statistical test was performed to determine whether the estimated PA realization rates were significantly different from the rest of the state. In both cases, the results indicated that the differences were significant, with 93% confidence for both National Grid and Columbia Gas. According to the Protocol this statistical difference, along with the sample size and precision level, would indicate that National Grid and Columbia Gas should both use their own realization rates and the other PAs should use the statewide result.

Table 4-5: Summary of Custom Gas Results by Program Administrator

Program Administrator	Annual Therms	Program Administrator	Annual Therms
Columbia Gas		**National Grid**	
Total tracking savings	700,742	Total tracking savings	1,079,572
Total measured savings	762,835	Total measured savings	811.829
Realization rate	108.9%	Realization rate	75.2%
Relative precision at 80% confidence	±19.1%	Relative precision at 80% confidence	±13.2%
Error bound at 80% confidence	146,065	Error bound at 80% confidence	106,997
Sample size	19	Sample size	15
Error ratio	3.68	Error ratio	0.56
Berkshire Gas		**NSTAR**	
Total tracking savings	10,110	Total tracking savings	188,112
Total measured savings	2,041	Total measured savings	159,617
Realization rate	20.2%	Realization rate	84.9%
Relative precision at 80% confidence	±21.2%	Relative precision at 80% confidence	±29.2%
Error bound at 80% confidence	432	Error bound at 80% confidence	46,568
Sample size	2	Sample size	7
Error ratio	0.58	Error ratio	0.59

Table 4-6: National Grid and Columbia Gas Individual Results vs. Rest of State

Program Administrator Group	Annual Therms	Program Administrator Group	Annual Therms
NGRID		**Columbia Gas**	
Total tracking savings	1,079,572	Total tracking savings	700,742
Total measured savings	811,829	Total measured savings	762,835
Realization rate	75.2%	Realization rate	108.9%
Relative precision at 80% confidence	±13.2%	Relative precision at 80% confidence	±19.1%
Error bound at 80% confidence	106,997	Error bound at 80% confidence	146,065
Sample size	15	Sample size	19
Error ratio	0.56	Error ratio	3.68
All Program Administrators, Excluding NGRID		**All Program Administrators, Excluding Columbia Gas**	
Total tracking savings	902,869	Total tracking savings	1,281,699
Total measured savings	919,483	Total measured savings	969,974
Realization rate	101.8%	Realization rate	75.7%
Relative precision at 80% confidence	±15.4%	Relative precision at 80% confidence	±12.3%
Error bound at 80% confidence	141,370	Error bound at 80% confidence	119,627
Sample size	28	Sample size	24
Error ratio	3.48	Error ratio	0.58

Analyses were also performed by measure category (across PAs). These results are presented in Table 4-7. Due to the fact the assignment of projects to measure categories was based on many assumptions, and the ERS findings that eight of fifty-four projects reviewed had been mis-classified, these results are provided for general information purposes only. It appears that the realization rates for boilers and insulation are low, while those for non-boiler heating and "other" were much higher. The realization rate for the controls category is high due the Outlier, a single site with an extremely high value. Even though this Outlier's weight has been adjusted, it is still influential.

Table 4-7: Results by Measure Category

Measure	Annual Therms
Boilers	
Total tracking savings	408,746
Total measured savings	176,620
Realization rate	43.2%
Relative precision at 80% confidence	±29.4%
Error bound at 80% confidence	51,897
Sample size	9
Error ratio	0.84
Controls	
Total tracking savings	460,645
Total measured savings	570,889
Realization rate	123.9%
Relative precision at 80% confidence	±21.6%
Error bound at 80% confidence	123,430
Sample size	9
Error ratio	3.26
Insulation	
Total tracking savings	221,256
Total measured savings	113,240
Realization rate	51.2%
Relative precision at 80% confidence	±9.8%
Error bound at 80% confidence	11,057
Sample size	9
Error ratio	0. 27

Measure	Annual Therms
Non-boiler heating	
Total tracking savings	408,371
Total measured savings	383,469
Realization rate	93.9%
Relative precision at 80% confidence	±15.7%
Error bound at 80% confidence	60,112
Sample size	9
Error ratio	0.43
Other	
Total tracking savings	483,423
Total measured savings	539,002
Realization rate	111.5%
Relative precision at 80% confidence	±11.7%
Error bound at 80% confidence	63,119
Sample size	7
Error ratio	0.44

5. Tracking Savings Estimation Methods Review

The section presents findings related to the PA methods of estimating savings and the documentation of those savings. More detailed descriptions and tables are presented in Appendix B.

5.1 Discrepancy Analysis

Each of the sites 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 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. Note that the Outlier site is excluded from this analysis because it swamps the variations analysis.

Appendix B discusses each of the categories in more detail and provides additional analysis of how particular sites influenced the outcomes in these categories. The discrepancy categories and highlights of each follows:

- **Tracking.** Represents those sites where the tracking data references an estimate of measure savings that was not supported by the project files. The source of the discrepancy might be clerical error, or it could be that subsequent analysis was not included in the final project folder.

- **Efficiency and Operational Baseline.** Occur when the evaluator does not agree with the applicant's baseline. A summary of sites with this discrepancy are listed in Table 5-1.

Table 5-1 Baseline Efficiency and Operational Discrepancies by Site

Site ID	PA	Sample Measure Description	Tracking Savings-MMBtu	RR%	Baseline Efficiency	Operation Baseline	Description
31	Columbia Gas	Boilers	172	25%	-52%		End of useful life
61	National Grid	Boilers	873	27%	-5%		Linkage controls, baseline efficiency change
64	Columbia Gas	Boilers	959	25%	-24%		End of useful life
71	National Grid	Boilers	5695	98%	-6%		Linkage controls, baseline efficiency change
93	Columbia Gas	Controls	2080	39%		-26%	EMS, baseline operation was more efficient than claimed in application
96	National Grid	Controls	3759	9%		-91%	Most EMS strategies required by code, except OSS and 365 Day Schedule
150	Columbia Gas	Insulation	90	37%	6%		Adjusted baseline model of windows to code
163	Columbia Gas	Insulation	117	39%	6%		Adjusted baseline model of windows to code
179	Berkshire Gas	Insulation	207	18%	-77%		End of useful life, re-roofing
256	NSTAR	Non-boiler heating	274	10%	-23%		End of useful life
278	NSTAR	Non-boiler heating	2120	43%	-37%		End of useful life

The most common baseline discrepancy occurred with an end-of-life equipment replacement measure where the applicant referenced the old equipment as the baseline. In the case of a boiler, domestic hot water, or roof insulation replacement, the owner is required to meet building code. However, it is also possible that an owner could be motivated to replace an operable older unit to save energy and then the baseline should be the old equipment even though a replacement is required to meet code. The evaluator interviewed the site contact concerning the functioning of the old equipment to ascertain the appropriate baseline. In those cases where the site contact stated the old equipment was at the end of life or was not functioning properly, code was selected as the baseline.

The operational baseline applies in the case of an EMS where the applicant claims a baseline operation that was not supported by the data.

- **Final Efficiency.** Captures discrepancies between the *installed equipment efficiency* projected by the applicant and that determined by the evaluator. This deviation applies to boilers, domestic hot water equipment, boiler control, and heat recovery equipment where the intent of the measure was to improve efficiency.

- **Operational.** Captures the operation of the post-installed equipment. This discrepancy captures, for example, the actual hours of equipment operation or the actual cfm of an exhaust system.

- **Non-Discernable.** Represents where the source of the discrepancy could not be pinpointed. For example, one site utilized a deemed percent savings of billed usage as the basis of the tracking estimate, while the evaluator used a bin analysis. It wasn't possible to attribute the difference in the results to any particular factor. Most of the sites had at least a portion of the total discrepancy assigned to this category.

- **Non-Operable.** Represents those sites where the installed equipment was not operating in the manner intended. It is possible that additional commissioning procedures at the PA inspection could have identified some of these conditions.

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 is 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 (in therms) to provide an estimate of the value of the discrepancy in the population. Estimates of program discrepancy impacts are provided in Table 5-2.

Table 5-2 Discrepancy Analysis Summary

	MMBtu	Weighted Net Pct	Sites Affected
Tracking error	(5,745)	-3.2%	6
Gas billing error	(3,067)	-1.7%	2
Quantity	315	0.2%	5
Baseline efficiency	(6,725)	-3.8%	9
Final efficiency	(5,448)	-3.1%	14
Operational baseline	(7,903)	-4.5%	2
Operational	(5,804)	-3.3%	13
Indiscernible	(3,670)	-2.1%	36
Non-operable	(10,581)	-6.0%	4
Other	(331)	-0.2%	2
Total	(48,958)	-27.7%	42

Two companion graphs are provided. Figure 5-1 graphs the sum of the absolute value of the discrepancies and the number of sites with the discrepancy and provides a view of the total variance of the factor and also the number of sites impacted.

Figure 5-1 Absolute Weighted Discrepancy

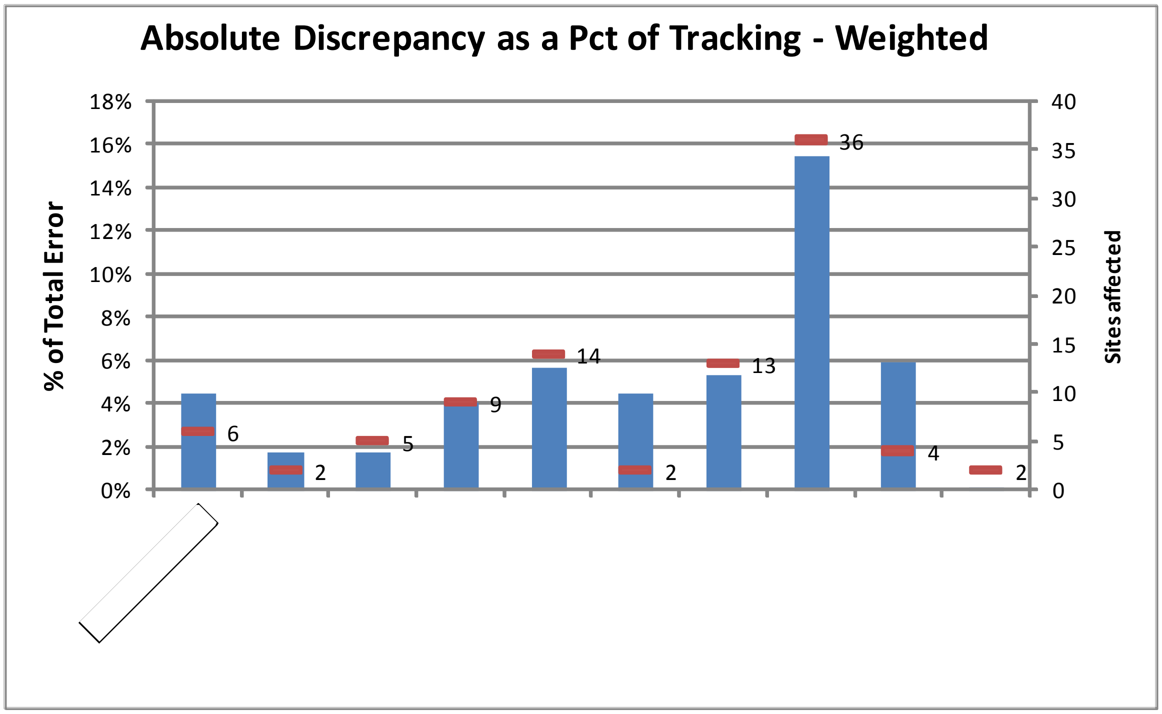
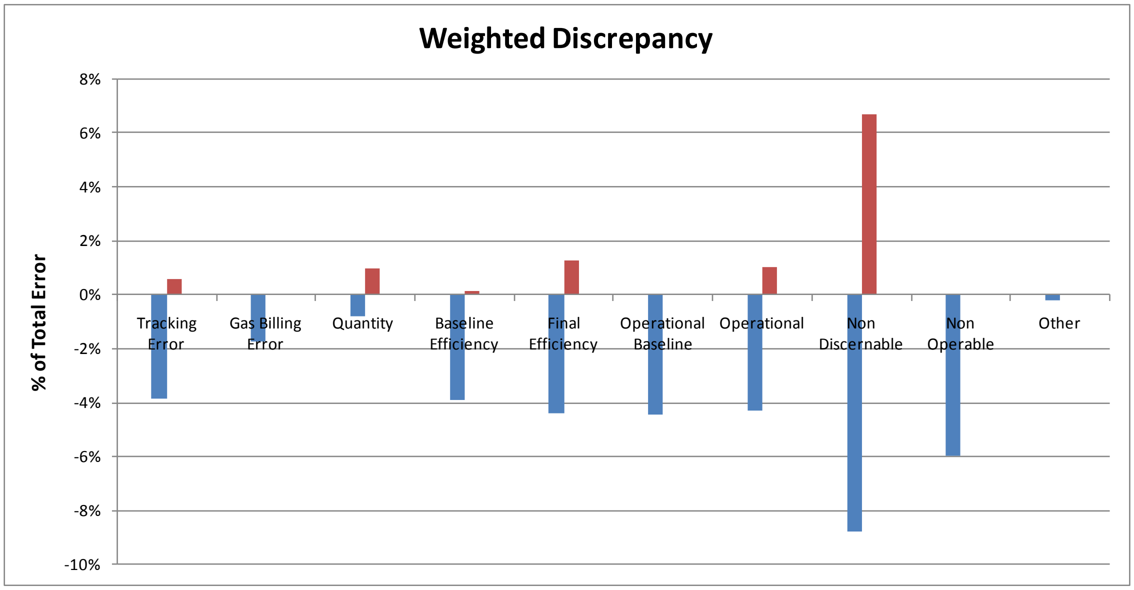


Figure 5-2 presents both the negative and positive component of the discrepancy. If the discrepancy was unbiased and random, the columns would be equally negative and positive, similar to the Non-discernable category. The baseline efficiency category is, however,biased, showing that tracking savings tend to over estimate savings due to incorrect baseline definitions. Most of the discrepancies tend to overestimate savings.

Figure 5-2 Weighted Contribution of Discrepancy



5.2 Tracking Estimation Methods and Project File Contents

A variety of methods were used to estimate savings ranging from simple equations through complex building simulation models. In some cases, it was not possible to discern the calculation method. In-house or proprietary models cannot be reviewed for methodology. While in some cases the equation methods produced good results, the equations often were little more than average estimates of savings and did not reflect the particulars of the site. Table 5-3 tabulates the analytical methods employed by applicants.

Table 5-3 Savings Estimation Methodology

PA	Building Simulation	Bin	Equation	In-house model	Unknown
Berkshire Gas			1		1
Columbia Gas	6	5	2	6	
National Grid	5	4	4		3
NSTAR		1	5		1
Grand Total	11	10	12	6	5

Project files ideally contain electronic copies of all the PA application paperwork and emails, applicant technical assistance (TA) studies, and native form analysis spreadsheets and models.

The evaluator also contacted both the PA and vendor that provided any customer services for this information as well. TA studies dated as far back as 2006. A summary of the project file content is shown in Table 5-5-4.

Table 5-5-4 Project File Content

	Billing History and Bill			Native Form		
PA	In the TA Study	New Const	Not Included	Not Avail	Provided	Not required
Berkshire Gas			2			2
Columbia Gas	4		15	9	8	2
National Grid	5	3	8	5	4	7
NSTAR	3		4		1	6
Grand Total	**12**	**3**	**29**	**14**	**13**	**17**

The PA realization rates are most directly impacted by a lack of paperwork as follows:

- Savings are assumed to be zero if there is insufficient documentation to assess the site. The EEAC Consultant has made it clear that sites should not be given a positive benefit of the doubt if the paperwork is poor. This did not occur in this study.

- When trying to determine if an equipment replacement measure is motivated primarily by energy reductions. The documentation must make this case abundantly clear; otherwise, code will be used as baseline.

- Measures where the baseline depends upon conditions that cannot be observed at a later date, such as steam trap failures or steam valve failures. This is discussed further in a subsequent section.

6. Conclusions and Recommendations

Overall, the Custom Gas program appears to be successfully providing energy savings in the State of Massachusetts. Below are major findings and recommendations that apply statewide.

6.1 Realization Rates

In November 2010, the Protocol was developed for applying realization rates to 2010 tracking savings for annual reporting purposes, as shown in the memo included in Appendix A. Subsequent to the memo, it was agreed that the results of the study would not be directly used for 2010 annual reporting purposes, since they would not be available in time to be used for this purpose.

Realization rates will be used, however, for program year 2012 mid-term modification planning and maybe be used for planning for the second three-year plan, 2011 reporting and succeeding program year reporting, unless replaced by results from a subsequent study.

While the Protocol called for estimating results by PA and measure category, it was expected that the achieved precision levels would be relatively “weak” for any particular measure category or PA. The Protocol called for the consideration of results at the following levels:

The study produced results for Level 1 that are reliable (±10.8%). For Level 2, the precision levels for National Grid and Columbia Gas were ±13.2% and ±19.1% respectively, which are reasonably good. The differences between each of these PAs and the corresponding “All Other PA” results were found to be statistically significant. Thus, the criteria defined in the Protocol for recommending that National Grid and Columbia Gas apply their own realization rates were met.

The results do not support application by measure, although on first glance, the statistical outcomes may indicate otherwise. The assignment of projects to measure categories was primarily based on a the description in the project title field; ERS found that eight of fifty-four projects reviewed had been misclassified. These results indicate interesting performance trends, but are not considered reliable for planning.

The PAs may also wish to consider whether to exclude the outlier site when planning. While the outlier impact is included in the retrospective analysis and realization rates presented in Section 4, the likelihood of another high gas usage site with grossly underestimated savings occuring in the future is low.

6.2 Savings Estimation Improvements

The evaluation team has reviewed project files, conducted detailed analysis of the information provided in the files, and used a discrepancy analysis to make the recommendations of this section.

6.2.1 Savings Estimation Procedures and Initial Screening

Major capital projects should be treated as equipment replacements. Boiler and DHW heater and addition of insulation upon re-roofing should be treated as equipment replacement opportunities. Savings should be calculated using the incremental costs and savings above a code equivalent replacement, since, typically, customers consider these measures only when at or near the end of life of the equipment.

Review measures and categorize correctly. It is recommended that all measures be reviewed to confirm that the measure is correctly categorized. This consistency is required before measure-level realization rates can be meaningfully applied. Creating a consistent categorization of measures will require sometime for development and dissemination of measure definitions, PA tracking system modification to support the categories, and training for the staff reviewing files.

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.

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.

6.2.2 Documentation Procedures

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 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. It is possible savings could be discounted to zero if there is not sufficient information to evaluate the site.

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.

6.2.3 Post Implementation Procedures

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.

6.3 Future Impact Issues

The outcomes of this study are useful for identifying future evaluation activities.

6.3.1 Implications for Future Sample Sizes

One of the greatest benefits of this study is the development of information on the site-to-site variability of savings to inform future sample designs. During the design of this study, an error ratio of 0.80 was assumed for annual therms. This was chosen to be conservative and turned out to be fairly accurate. Based on the findings of this study, including the single influential site, the overall error ratio was higher than anticipated at 1.93. However, if this site is excluded, which may be appropriate for future study planning, the error ratio was 0.68.

6.3.2 Value of Accelerated Replacement

The evaluators found that many of the boiler replacement and other large capital projects were end-of-life projects that required code as the baseline. However, it is likely that in some cases the energy efficiency program accelerated the replacement of the equipment. If this period of time could be identified, additional savings could be assigned to the project equal to the difference between the replaced equipment efficiency and code multiplied by the years of acceleration. This can significantly add to the overall savings.

6.3.3 Pre-Metering of Select Measures

Some measure categories are difficult to evaluate after the installation without sufficient pre-installation data that characterizes the existing equipment operation, including steam trap and control valve failures. This is problematic with thermostatic radiator valve (TRV) measures and measures relying on designs that eliminate steam traps, since the savings depends upon the existing equipment's operational status (number of failed traps or valves). If the implementation of the measure occurs in isolation of other measures and the savings are large enough, it may be possible to use billing analysis to reliably determine savings.

The PAs may wish to consider evaluation activities consisting of pre-inspection or pre-metering for sites where significant savings are expected from these types of measures.

Appendix A: MA-LCEIC Custom Gas Impact Evaluation Application of Results, November 23, 2010

Appendix B: Review of Tracking Savings Estimation Methods

Appendix C: Site Reports

Appendix A: MA-LCEIC Custom Gas Impact Evaluation

Application of Results

November 23, 2010

A-1 Background

The MA-LCIEC Custom Gas Impact Evaluation project is underway, with initial data collection for sample projects in progress. The purpose of this document is to describe the groups for which results will be estimated, tested, and applied to 2010 program year evaluations. While realization rates can and will be developed for many levels of aggregation, the criteria for determining which will be used for each Sponsor and measure category have not been finalized. It is important to agree upon the decision rules for applying study results in advance of the actual calculations, to ensure fair coverage for all Sponsors. By applying the decision rules described below, only those results that are based on sample sizes that are adequate to ensure reasonable statistical validity will be used.

A-1.1 Sample Design and Selection

The sample design and selection for the Custom Gas Impact Evaluation is complete and includes forty-seven sites that will be metered and evaluated to estimate savings realization rates. One design objective was to support the estimation of realization rates for therms saved across a number of efficiency measure categories including:

- Boilers
- Controls
- Insulation
- Non-Boiler Heating
- Other

Another objective was to estimate statistically valid realization rates for each Sponsor. In fact, the number of sites budgeted for this project is insufficient to achieve both objectives. If the matrix of domains to be evaluated has thirty cells (five measure categories times six Sponsors), the number of sites required to achieve 80/10 precision targets with coverage in every cell would have been over five times the current budget.

The approved sample allocation by Sponsor and measure is included as Table 1. There are a few minor adjustments that will be required to account for the fact that at least two of the projects selected have been determined to be prescriptive measures under current definitions.

Table 1: Sample Allocation by Sponsor

Sponsor	Boilers	Controls	Insulation	Non-Boiler Heating	Other	Total
N-GRID	5	4	1	1	4	15
NSTAR	1	1	1	2	1	6
NE Gas	0	0	0	1	1	2
Berkshire Gas	0	1	1	0	0	2
Bay State Gas	3	3	6	5	2	19
Unitil	0	1	0	1	1	3
Total	9	10	9	10	9	47

A-1.2 Recommended Application of the Results

We support the notion of obtaining foundational information across all five measure categories and all Sponsors. However, it is likely that achieved precision levels will be relatively “weak” for any particular measure category or Sponsor. Once the data has been collected from the sample sites, realization rates and their associated precision levels will be estimated at the following levels of aggregation:

- Level 1: State-wide realization rate, across all measure categories
- Level 2: Realization rate by Sponsor, across all measure categories
    - Calculate separate realization rates for Sponsors with at least ten projects (National Grid and Bay State Gas).
    - Aggregate the Sponsors with less than ten projects each into one group.
    - Calculate two “all other sponsor” realization rates by first excluding National Grid and then excluding Bay State Gas.

- Level 3: For the Sponsors with five or more sample projects in a given measure category, calculate realization rate by Sponsor and measure category.
- Level 4: State-wide realization rate for each measure category.

It is possible that only Level 1 will produce results that are reasonably reliable (better than ±20%). The relative precision of results at Level 2 and Level 3 will be evaluated to determine whether they are sufficiently precise and significantly different to be used individually. For Level 2, if the Sponsor-specific realization rate (across all measures) attains a minimal precision threshold, e.g., ±35%, and if the Sponsor-specific realization rate is found to be statistically different than the "all other sponsor" realization rate, e.g., at the 20% level of significance, then the Sponsor should be required to use their Sponsor-specific realization rate in place of the State-wide realization rate. For National Grid's Level 2 test, the all-other-sponsor rate would be calculated based on all Sponsors except National Grid. Similarly for Bay State Gas's Level 2 test, the all-other-sponsor rate would include all Sponsors except Columbia Gas. For Level 3, the same two criteria would be applied by Sponsor and measure category.

The estimation of Level 4 realization rates will probably not produce results that are reliable. However, the knowledge gained by identifying areas where savings calculations are particularly inaccurate will help direct future research. The error ratios that are determined from the sample data will allow for the more solid calculations of sample sizes required to achieve desired precision levels in the next phase of evaluating Custom Gas measures. In the long term, the precision of estimated realization rates can be improved by addressing the methods for calculating savings for the most volatile measures.

Appendix B: Review of Tracking Savings Estimation Methods

The section presents findings related to the PA methods of estimating savings and the documentation of those savings. The first section provides additional details of the discrepancy analysis. The second section presents more detailed observations and recommendations regarding PA customer tracking estimation processes.

B-1 Discrepancy Analysis

The key result of the study is the evaluated typical annual gas savings and the site realization rate, which are shown in **Error! Reference source not found.** in the body of the report. Additional quantitative analysis was conducted to determine why the evaluator's savings differed from that of the applicant's. Each of the sites was reviewed to determine the factors impacted the savings and judgments made to categorize them. Calculations were conducted to determine the contribution of that factor to the final deviations in savings from tracking

Note that the process of categorizing and quantifying discrepancies required judgment and is not formulaic. However, the analysis does identify patterns that indicate where the estimation process could be improved.

B-1.1 Tracking and Gas Billing Discrepancies

Sites with tracking and gas billing discrepancies are summarized in Table 1.

Table 1: Sites with Tracking and Gas Billing Errors

Site ID	PA	Tracking Savings-MMBtu	RR%	Tracking Error	Gas Billing Error	Notes
93	Columbia Gas	2080	39%		-27%	Referenced another school's bills
202	Columbia Gas	18	206%	284%		Unknown reason
40	Columbia Gas	340	30%	-2%	-37%	Referenced another school's bills
256	NSTAR	274	10%	-25%		Vendor calculations were different
66	National Grid	1929	21%	-41%		Increased savings with more boilers, but same output
44	National Grid	474	37%	-52%		Savings not updated with new calcs

The tracking category represents those sites where the tracking data references an estimate of measure savings that was not supported by the project files. The source of the deviation might be clerical error or a misread of the analysis, or it could be that tracking estimates were not updated with subsequent analysis.

The gas billing category represents those sites where the applicant calculations clearly referenced the wrong gas bills leading to miscalculated savings. At some sites, the gas usage referenced in a calculation was not directly linked to a particular billing period or adjustment process (such as weather normalization). The usage may have looked suspicious, but that site was not included in the gas-billing-discrepancy count.

B-1.2 Efficiency and Operational Baseline Discrepancies

The baseline deviation occurs when the evaluator does not agree with the applicant's baseline. Table 2 summarizes the sites with baseline deviations. (This is a duplicate of **Error! Reference source not found.**).

Table 2: Sites with Efficiency and Operational Baseline Discrepancies

Site ID	PA	Sample Measure Description	Tracking Savings-MMBtu	Eval Elec	Tracking MMBTU	Baseline Efficiency	Operation Baseline	Description
31	Columbia Gas	Boilers	172	25%	172	-52%		End of useful life
61	National Grid	Boilers	873	27%	873	-5%		Linkage controls, baseline efficiency change
64	Columbia Gas	Boilers	959	25%	959	-24%		End of useful life
71	National Grid	Boilers	5695	98%	5,695	-6%		Linkage controls, baseline efficiency change
93	Columbia Gas	Controls	2080	39%	2,080		-26%	EMS, baseline operation was more efficient than claimed in application
96	National Grid	Controls	3759	9%	3,759		-91%	Most EMS strategies required by code, exce[t OSS and 365 Day Schedule
150	Columbia Gas	Insulation	90	37%	90	6%		Adjusted baseline model of windows to code
163	Columbia Gas	Insulation	117	39%	117	6%		Adjusted baseline model of windows to code
179	Berkshire Gas	Insulation	207	18%	207	-77%		End of useful life, re-roofing
256	NSTAR	Non-boiler heating	274	10%	274	-23%		End of useful life
278	NSTAR	Non-boiler heating	2120	43%	2,120	-37%		End of useful life

The most common baseline discrepancy occurred with an end-of-life equipment replacement measure where the applicant referenced the old equipment as the baseline. In the case of a boiler, domestic hot water, or roof replacement, the owner is required to meet building code. However, it is also possible that an owner could be motivated to replace an operable older unit to save energy and then the baseline should be the old equipment even though a replacement is required to meet code. The evaluator interviewed the site contact concerning the functioning of the old equipment to ascertain the appropriate baseline. In those cases where the site contact stated the old equipment was at the end of life or was not functioning properly, code was selected as the baseline.

In the case of the two linkage control measures, the baseline efficiency of the prior system with linkage controls was not supported by the data and was adjusted accordingly. Two sites with window replacements had modeled the baseline windows slightly better than code.

The operational baseline category applies in the case of an EMS where the applicant claims a baseline operation that differed from the evaluator. In one case, the EMS was part of a new construction project and most of the savings strategies are required by code and therefore were discounted. At a second site with EMS measures, the baseline assumed continuous operation of equipment while the site contact stated manual control was in place. The manual control operation was supported by a billing analysis as well.

B-1.3 Quantity and Inoperable Discrepancies

These two categories capture the sites where the installed quantity was in disagreement with the application or where the installed equipment was inoperable. In theory, this category could partially be improved through inspections and commissioning procedures. The sites are summarized in Table 3.

Table 3: Sites with Quantity and Inoperable Discrepancies

Site ID	PA	Building Type	Sample Measure Description	Tracking Savings-MMBtu	RR%	Quantity	Non Operable	Notes
195	National Grid	Nursing home	Insulation	733	57%	-36%		Difference in insulated area
147	Columbia Gas	Office	Insulation	86	72%	-8%		Difference in insulated area
163	Columbia Gas	Public housing	Insulation	117	39%	31%		Difference in window area
171	Columbia Gas	Public housing	Insulation	142	37%	40%		Difference in window area
150	Columbia Gas	Public housing	Insulation	90	37%	47%		Difference in window area
68	National Grid	Lodging - large hotel	Boilers	1997	0%		-100%	Controls did not yield savings
85	Columbia Gas	Mixed use, very large space	Controls	558	0%		-100%	Controls and strategy not installed
305	National Grid	Office - gym	Other	222	0%		-100%	No airflow through heat wheel
332	Columbia Gas	High school	Other	1688	31%		-69%	Setpoints incorrect, dampers not working

B-1.4 Final Equipment Efficiency Discrepancy

This category captures deviations in the installed equipment efficiency projected by the applicant and that determined by the evaluator. This discrepancy applies to boiler, domestic hot water equipment, boiler control, and heat recovery equipment where the intent of the measure was to improve efficiency.

The evaluator used seasonally adjusted combustion efficiency in most cases for computation of this metric, while the applicant reference value is usually maximum combustion efficiency. The applicant paperwork typically represented this maximum combustion efficiency as a seasonal efficiency value (at least by inference), and usually the algorithms were not specific enough to discern if a more sophisticated approach had been taken. Table 4 provides a list of sites with final equipment discrepancies.

Table 4: Sites with Final Equipment Discrepancies

Site ID	PA	Building Type	Sample Measure Description	Tracking Savings-MMBtu	RR%	Final Efficiency
61	National Grid	University dormitory	Boilers	873	27%	-48%
256	NSTAR	Dry cleaners and laundry	Non-boiler heating	274	10%	-22%
66	National Grid	Public housing	Boilers	1929	21%	-20%
44	National Grid	Public housing	Boilers	474	37%	-16%
64	Columbia Gas	Elementary school	Boilers	959	25%	-16%
266	NSTAR	Apartment building	Non-boiler heating	494	23%	-12%
40	Columbia Gas	Elementary school	Boilers	340	30%	-11%
31	Columbia Gas	Apartment building	Boilers	172	25%	-9%
278	NSTAR	Apartment building	Non-boiler heating	2120	43%	-4%
282	National Grid	Pharmaceutical manufacturer	Non-boiler heating	7199	99%	-1%
71	National Grid	Industrial food processing	Boilers	5695	98%	18%
62	NSTAR	Office building	Boilers	920	73%	29%

B-1.5 Operational Differences

This error captures deviations in the projected hours of operation of the equipment or other metric of use (including total cfm). This is sometimes the most difficult figure for an applicant to project, although with gas measures, such as a boiler replacement, using the gas bills as a starting point can remove some of this uncertainty. Table 5 provides sites with operational differences.

Table 5: Sites with Operational Differences

Site ID	PA	Building Type	Sample Measure Description	Tracking Savings-MMBtu	RR%	Operation	Notes
150	Columbia Gas	Public housing	Insulation	90	37%	-78%	Calibration to bills, lower usage
171	Columbia Gas	Public housing	Insulation	142	37%	-71%	Calibration to bills, lower usage
163	Columbia Gas	Public housing	Insulation	117	39%	-68%	Calibration to bills, lower usage
266	NSTAR	Apartment building	Non-boiler heatin	494	23%	-45%	Metered water usage is less
99	National Grid	Paint manufacturing	Controls	7331	65%	-29%	Impacted hours are fewer, less CFM
71	National Grid	Industrial food processing	Boilers	5695	98%	-13%	Facility usage is down
256	NSTAR	Dry cleaners and laundry	Non-boiler heatin	274	10%	-11%	Metered water usage is less
235	Columbia Gas	Public housing	Non-boiler heatin	126	61%	-5%	Calibration to bills, lower usage
147	Columbia Gas	Office	Insulation	86	72%	-5%	Calibration to bills, lower usage
339	National Grid	Industrial food processing	Other	8800	127%	12%	Evaluator used alternate orifice characteristics
324	Columbia Gas	Small industrial	Other	645	300%	48%	Impacted hours are 3x estimated by applicant
89	Columbia Gas	Mixed use, very large space	Controls	794	3412%	3312%	Applicant was conservative; high vacancy rate driving savings

B-2 Custom Process Findings and Recommendations

The detailed review of project file information and vendor estimates suggest some practical improvements that the PAs can make to the custom application process. The suggestions are organized as follows:

- Treatment of equipment replacement measures
- Project file content
- Use of gas bills

B-2.1 Treatment of Equipment Replacement Measures

Retrofit measures are generally viewed as measures that occur at the discretion of the owner primarily motivated to save energy. Retrofit measures are screened at the full cost of the install with savings calculated as the difference in usage between the previous existing equipment and the new equipment. Equipment replacement measures are generally viewed as time-sensitive measures that an owner must complete. Equipment replacements are screened using the incremental cost of the measure with savings calculated as the difference in usage between code and the new equipment.

The evaluator concludes from a review of these measures that boiler, domestic hot water, and membrane roof insulation should be treated as equipment replacement opportunities in the future. Savings should be calculated using the incremental costs and savings above a code equivalent replacement. While equipment replacement is often considered a “lost opportunity” measure, the program implementers may wish to position the measure in the retrofit portfolio, if it represents a successful delivery model.

Where the PAs believe a site is truly a retrofit application because the customer is making the decision primarily based on energy efficiency, the file documentation becomes particularly critical and should:

- Document the financial package. Include installed cost, incentives, and projected annual dollar savings. Document the customer’s financial screening criteria.
- Document the satisfactory operation of the existing equipment. This might include photos (for shell measures), identification of other similar units that were not replaced in the facility, use of maintenance records or other means.
- Document the make and model of the previous equipment.

B-2.2 Project File Contents

Project files ideally contain electronic copies of all the PA application paperwork and emails, applicant technical assistance (TA) studies, and native form analysis spreadsheets and models. The evaluator also contacted the PA and the vendor that provided any customer services for this information as well. TA studies dated as far back as 2006.

The PA realization rates are most directly impacted by a lack of paperwork as follows:

- Savings are assumed to be zero if there is insufficient documentation to assess the site. The EEAC Consultant has made it clear that sites should not be given a positive benefit of the doubt if the paperwork is poor.
- When an equipment replacement measure is motivated primarily by energy reductions. The documentation must make this case abundantly clear; otherwise, code will be used as baseline.
- Measures where the baseline depends upon conditions that cannot be observed at a later date, such as steam trap failures or steam valve failures. This is discussed further in a subsequent section.

It was intended that the evaluator use the same algorithms as the applicant; however, this was often not possible because the project files did not include the applicant's spreadsheet or enough information to replicate the methodology. It is likely that the use of the applicant's algorithm would *not* have resulted in significantly different savings estimates, except as noted above, because most of the deviations were traced to factors such as final efficiencies and baseline changes, which would have impacted the applicant's algorithm as well.

Better documentation would provide better insights into the sources of deviations, which might lead to better program implementation.

As evaluators, native building simulation models and spreadsheets were sorely missed. Because this information was rarely available, the team had to start from scratch, which adds to the cost of the studies.

B-2.3 Document and Use Customer Billing

Accurate customer gas billing is the most important single piece of information in the project folder. The PAs should implement administrative practices that capture the customer bill and use them to verify the measure savings as part of the normal technical review.

Billing documentation should include:

- Copy of the gas bill with service address, customer name, and meter number. The site should be checked for more than one gas meter serving the site and/or the measure. This provides the PA reviewer with an opportunity to make sure the TA has referenced the correct gas usage for the appropriate location. It permits the evaluator to reliably locate the measure. This is particularly important for individually metered housing complexes, where building designations are ambiguous. About a third of the sites were multi-family units.

- Copy of the last 18 months of billing usage at the time of the TA study and also at the time of the incentive payment. Some measures are identified as part of the TA study but are implemented much later. The gas distribution companies keep limited data active in Customer Information Systems; therefore billing data from before implementation is not always available. If the billing usage is part of the project file, the PA can sanity check the savings, and the evaluator will always have pre-implementation billing.

- Document the exact period of measure implementation.

Each custom application should include a sanity check of the results using the customer billing data. It's possible this check could even be automated at the PA's discretion and could be refined by calculating base vs. weather-sensitive usage and use weather normalizing. The check is a simple calculation of the applicant savings divided by the gas usage with a cross-check of the reasonableness of that estimate. The intent of this checking process is to reduce obvious errors related to incorrect bills and to provide a quick screen of those sites where the applicant savings estimates need further scrutiny.

Finally, simulation models and bin models should be calibrated to the bills.