Existing Facilities Program & National Fuel's Non-Residential Rebate Program: Impact Evaluation (January 2012 - September 2013)

Final Report

November 2015

Prepared for:

The New York State
Energy Research and Development Authority

And

National Fuel Gas Distribution Corporation

Judeen Byrne
Emily Shusas
Project Managers

Prepared by:

ERS
120 Water Street, Suite 350
North Andover, MA 01845
Phone: 978-521-2550

Principal Investigators:
ERS
Cx Associates
West Hill Energy and Computing, Inc.

NOTICE

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

ABSTRACT

This report describes the impact evaluation of New York State Energy Research and Development Authority's (NYSERDA's) Existing Facilities Program (EFP) and National Fuel Gas Distribution Corporation's (NFGDC) Non-Residential Rebate Program (NRCIP). NYSERDA administers NFGDC's NRCIP.

The Impact Evaluation Team assessed the energy savings of the EFP and NRCIP for projects with measures completed between January 1, 2012, and September 30, 2013. The Impact Evaluation Team performed measurement-based engineering analysis on a sample of completed custom electric and custom and prequalified natural gas projects to quantify the evaluated gross energy savings by project. Prequalified electric was deemed based on prior evaluation results. For NYSERDA, the EFP electric realization rate is 1.01, and the natural gas realization rate is 0.64. For NFGDC, the natural gas realization rate is 0.66. The report also provides estimated savings and realization rate of 1.02 for EFP projects installed in 2010 and 2011 based on this study's research and that of the prior EFP impact evaluation. This evaluation did not include attribution.

Through project-level research, the Impact Evaluation Team concludes that a large range of project types and sizes with varying technologies and stake holders have been implemented.

This report presents the evaluated gross energy savings of the EFP and NRCIP; net impacts of free ridership and spillover were not in scope.

ACKNOWLEDGMENTS

This final report was prepared by the Impact Evaluation Team, led by ERS with Yogesh Patil as the project manager. ERS and Cx Associates conducted the on-site verification. West Hill Energy consulted with ERS on the sample design and aggregate analysis. The Impact Evaluation Team wishes to acknowledge the significant contribution of NYSERDA evaluation staff Judeen Byrne and Emily Shusas, NFGDC's evaluation staff Evan Crahen and Eric Meinl, the input of the NYSERDA and NFGDC Program staff, and also the valuable suggestions provided by Bill Saxonis of the DPS and its consultants, Ralph Prahl and Rick Ridge of TecMarket Works.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

SECTION 1: EXECUTIVE SUMMARY

This report describes the impact evaluation of projects that were installed between January 1, 2012 and September 30, 2013 for the New York State Energy Research and Development Authority's (NYSERDA) Existing Facilities Program (EFP) and the National Fuel Gas Distribution Corporation's (NFGDC) Non-Residential Rebate Program (NRCIP). Jointly, these programs are being referred to as the Program throughout this evaluation report. NYSERDA administers the NRCIP through the EFP.

The primary objective of this impact evaluation was to estimate the evaluated gross electric and natural gas savings. To best align with and provide actionable insights for NYSERDA's developing future program plans, at the time this study was scoped, the evaluation did not address gross savings for electric prequalified projects.

However, the program electric evaluation realization rate (RR) results include both custom and prequalified projects.

1.1 APPROACH

The Impact Evaluation Team estimated savings based on project-specific measurement and verification (M&V) performed on a statistically valid sample of 61 custom electric energy savings projects and 53 custom and prequalified natural gas savings projects, and used deemed savings from the prior EFP evaluation for electric prequalified projects. The overall level of engineering rigor was high, with engineers visiting 100% of the evaluated (i.e., sampled) projects to inspect and verify installation. Ninety-eight percent of the sample was subjected to site-specific logging of equipment, site-specific billing analysis or both to estimate savings. Reviews of RRs as a function of measure type, building type, geographic location, sponsor, size, and other factors were examined to identify trends and opportunities for improvement.

The Impact Evaluation Team also interpolated the Program's 2010–2011 electric RR using results of this evaluation and those of the prior impact evaluation. Additional information on this is provided in Sections 1.2 and 3.3.

1.2 RESULTS

Table 1-1 summarizes the results of the evaluation.

Table 1-1. Impact Evaluation Results

Parameter	Program-Reported Savings (1/1/12 – 9/30/13)[1]	Realization Rate	Evaluated Gross Savings	Relative Precision
NYSERDA				
Electric energy (MWh/yr)	221,473	1.01[1,2]	223,688[2]	N/A[3]
Electric demand (MW)	35.9	0.98[1]	35.2	N/A
Natural gas energy (MMBtu/yr)	81,448	0.64[4]	52,127	7.9%
NFGDC				
Natural gas (MMBtu/yr)	33,137	0.66[5]	21,870	26.2%

[1] The electric RRs are based on the weighted average of the custom electric measure evaluated RRs from this study's primary research and the prequalified electric evaluated RRs from the prior EF impact evaluation.

[2] Realization rate and evaluated savings are rounded.

[3] The relative precision associated with the custom electric project sample is 6.4%.

[4] Due to program changes that occurred in 2014, the Impact Evaluation Team recommends that NYSERDA use an RR of 0.63 as a prospective pro forma value for 2014 and later reporting, until the next evaluation is completed.

[5] Due to program changes that occurred in 2014, the Impact Evaluation Team recommends that NFGDC use an RR of 0.80 as a prospective pro forma value for 2014 and later reporting, until the next evaluation is completed.

Calculations for the NYSERDA 2010–2011 period produced an electric RR of 1.02.

1.3 FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

The Impact Evaluation Team offers one finding and two recommendations based on the impact evaluation research; these are based on the sampled projects which included custom electric projects, and both custom and prequalified gas projects.

Finding – No Action Required

Although the Impact Evaluation Team found that the deemed savings for the prequalified natural gas furnace and boiler measures used an incorrect basis for projects reported complete during the evaluation period and this inflated the savings estimates and reduced the RR. Program staff fixed the deemed savings tool in 2014. The Impact Evaluation Team reviewed the revised tool and found it to be accurate. No further action is recommended of Program staff. As footnoted in Table 1-1, to the extent an RR is to be applied prospectively or on a pro forma basis, use 63% for NYSERDA and 80% for NFGDC, since this systematic issue has been addressed. These values considered the adjustments to the RR due to revised tool.

Conclusions and Recommendations

1. The Program estimates custom electric energy savings well.

The results of the past two evaluations have resulted in RRs in close proximity to 1, relative precision values within the range of the sample design, and low error ratios. Program staff and their third-party Outreach and Technical Reviewers have demonstrated that they are capable of estimating project savings with high levels of engineering rigor.

Recommendation: This Program need not be subject to a traditional comprehensive retrospective gross impact evaluation of electric custom savings in the next cycle unless the Program undergoes substantial changes.

2. Prequalified gas applications tend to have the most variation in RRs.

Prequalified[1] gas projects were found to have the most variability between the reported and evaluated savings, which resulted in lower relative precision values. By analyzing the RR differences attributable to the prequalified projects, Program staff can isolate differences that can be fixed during the application review process. Pre-project characterization was found to be one of the most common reasons for differences. Program staff could start by focusing on capturing the pre-project information accurately, which would result in fewer differences in savings values.

Recommendation: Put procedures in place that will help take a closer look at some of the differences and associated inputs. This will subsequently improve the prequalified project savings estimates. Specifically, the two biggest causes of underestimation of prequalified reported compared to the evaluated savings were differences in equipment load profiles and differences in equipment operating hours. The program staff should be more attentive to the screening of prequalified projects as they tend to have the most variations.

.

[1] Measure reviewed for cost-effectiveness by the DPS and approved for offering within a program. Prequalified electric incentives encourage customers working on small-sized energy projects and equipment replacement projects to purchase and install more energy efficient measures. In some cases prequalified measures have deemed savings; in other cases the program assigns application-specific savings estimates.

SECTION 2: INTRODUCTION

This section presents a program description, the evaluation goals, and a summary of the previous evaluations.

2.1 PROGRAM DESCRIPTION

The EFP promotes energy efficiency and demand management. In 2011, demand management measures were transitioned to the Technology and Market Development Program (T&MD) portfolio of programs; consequently, demand management measures were not assessed in this evaluation.

EFP targets a range of customer sectors, including commercial and industrial businesses, health care facilities, universities and colleges, state and local governments, and mission-critical facilities such as data centers and communications facilities.

There are two types of EFP incentives: prequalified and performance-based.

1. Prequalified[2] electric and gas incentives encourage customers working on small-sized energy projects and equipment replacement projects to purchase and install more energy efficient measures. The measures available to qualifying customers include lighting, chillers, motors, variable frequency drives (VFDs), and heating, ventilation and air conditioning (HVAC)
2. Performance-based incentives are for customers or energy service companies (ESCOs) working on large-scale projects. The incentives are typically higher than those for prequalified projects, and the performance-based projects require an engineering analysis and are potentially subject to a program required measurement and verification (M&V). The various types of performance-based incentives encourage the implementation of projects that deliver verifiable annual electric and gas savings.

NFGDC's NRCIP is an equipment replacement program, modeled after a Vermont Gas Systems program that was cited by the ACEEE as an exemplary natural gas energy efficiency program. NRCIP, administered by NYSERDA through EFP, offers equipment replacement prequalified and customized rebate incentives to customers using less than 12,000 Mcf[3]/yr; the purpose is to promote the installation of high efficiency space heating, water heating, and process heating equipment. Customers are also eligible to receive customized rebates for non-equipment

[2] In some cases prequalified measures have deemed savings; in other cases the program assigns application-specific savings estimates.

[3] Mcf is 1,000 cubic feet.

replacement changes made to space heating, water heating, and process heating equipment, such as adding insulation to a process heating oven, or updating controls to a space heating boiler. These custom incentives are set on a case-by-case basis, based upon the estimated resulting gas energy savings; however, a technical engineering analysis must first be performed to confirm the energy savings.

2.1.1 Summary of Program-Reported Savings

This evaluation included projects with at least one measure installed between January 1, 2012 and September 30, 2013. Table 2-1 presents the program-reported savings for these projects.

Table 2-1. EF Program-Reported Savings (2012–2013)

Measure Type/Facility Type	# Projects with Completed Measures	Program Reported Annual Savings	Percentage of Installed Savings
NYSERDA Electric			
Prequalified	1,888	118,207	53%
Custom	494	103,266	47%
Electric savings total	2,382	221,473 (MWh)	100%
NYSERDA Natural Gas			
All natural gas projects	35	81,448 (MMBtu)	100%
NFGDC Natural Gas			
All natural gas projects	129	33,137 (MMBtu)	100%

2.2 EVALUATION OBJECTIVES

The primary objective of this impact evaluation was to estimate the evaluated electric and natural gas savings for the Program installed between January 1, 2012 and September 30, 2013. The primary data collection activity consisted of rigorous project-specific M&V of a statistically valid sample of natural gas and custom electric projects installed during this time period. The Impact Evaluation Team also estimated EFP electric RRs for 2010 – 2011 by interpolating based on the results of this and the prior EFP evaluation.

During the course of the current evaluation NYSERDA modified the work scope to exclude electric prequalified measures from the field M&V focusing on custom measures only, and to exclude net savings/attribution.

The New York Department of Public Service (DPS) is the oversight agency for all program evaluation activity. NYSERDA has provided the DPS with key documents for review and comment throughout the evaluation. This report complies with the savings-related requirements

listed in *New York Evaluation Plan Guidance for EEPS Program Administrators*[4], which was issued by the DPS and is intended to provide robust, timely, and transparent results.

The overall evaluation scope and objectives are identified in Table 2-2.

Table 2-2. Evaluation Scope and Objectives

Objectives	Outputs	Method Used
Evaluated gross electric energy and gas savings	Annualized first-year evaluated gross energy savings based on electric (kWh) and natural gas savings (therms) at the customer meter. Evaluate custom projects only for electric and both prequalified and custom projects for natural gas.	On-site measurement and verification for 114 projects using on-site logging, custom engineering assessments, billing analysis, and building simulation of a representative sample of Program participants.
Realization rate (RR)	Ratio of the sum of the weighted evaluated savings divided by the sum of the weighted program-reported savings	
Statistical validity	The sample design targets 8% precision level for Program energy savings at 90% confidence.	Stratified ratio estimation sample design

2.3 PREVIOUS EVALUATIONS

The Program was previously evaluated in 2010–2011 for electric measures installed from 2006 through 2009. The work included on-site M&V activities at a sample of 92 sites and enhanced participant and vendor surveys to establish free ridership (FR) and spillover (SO) for those same sites. Natural gas savings were not evaluated since they were not being reported by NYSERDA.

The primary findings of the 2006–2009 EFP impact evaluation are summarized in Table 2-3.

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

Table 2-3. NYSERDA Prior Evaluation[1] for Projects Installed from 2006 through June 2009

	Energy Efficiency (MWh/yr)	Coincident Peak Demand Savings (MW)
Program-reported savings	577,787	116
Realization rate	1.03	0.81
Evaluated gross savings	595,121	94
Net-to-gross ratio calculation[2,3],	= 1 – 0.31 + 0.12 + 0.32 + 0.15	= 1 – 0.31 + 0.12 + 0.32 + 0.15
Net-to-gross ratio	1.28	1.28
Total evaluated net savings	761,755	120
RR precision	9.8%	8.0%

[1]Results extracted from "NYSERDA 2006–2009 Existing Facilities Program Impact Evaluation Report – Final," available at: http://www.nyserda.ny.gov/-/media/Files/Publications/PPSER/Program-Evaluation/2012ContractorReports/2012-EFP-Impact-Report-with-Appendices.pdf

[2]Net-to-gross ratio = 1 – Free ridership + Inside spillover + Outside spillover + Nonparticipant spillover

[3]Free ridership (FR) refers to Program participants who would have implemented the Program measure or practice in the absence of the Program. Spillover (SO) refers to reductions in energy consumption that could result from: a) additional energy efficiency actions that Program participants take outside the Program as a result of having participated; b) changes in the array of energy-using equipment that manufacturers, dealers, and contractors offer all customers as a result of Program availability; and c) changes in the energy use of nonparticipants as a result of NYSERDA programs. Additional energy efficiency actions that Program participants take inside the dwelling or facility served by the Program are referred to as inside spillover (ISO), while actions participants take or influence at other facilities not directly served by the Program are considered outside spillover (OSO). Savings from a nonparticipant who adopted a particular efficiency measure or practice as a result of the Program but does not participate in the Program are referred to as nonparticipant spillover (NPSO).

SECTION 3: METHODS

The scope of work was to determine the evaluated gross savings of both EFP and NRCIP. The methods used to research and evaluate the savings are discussed in the following sections.

3.1 EVALUATION METHODS

A critical component of the evaluation was the development of rigorous estimates of the project RRs for program-reported natural gas savings, and electric custom savings for EFP, which includes verifying the installation of efficiency measures and the generation of an independent savings analysis. On-site verification of savings included site M&V and site survey work on a representative sample of participating custom electric and natural gas projects. The project level savings are then weighted to calculate the program-level RR.

The program level RR is applied to the program-reported savings, resulting in the evaluated gross savings estimates. RRs represent an adjustment to the program-reported savings, upward or downward, to account for differences between the evaluated gross savings and program-reported savings.

$$RR = \frac{Savings_{Evaluated}}{Savings_{Reported}}$$

where,

RR = Realization Rate

$Savings_{Evaluated}$ = Savings as per measurement and verification (M&V) evaluation

$Savings_{Reported}$ = Savings as reported by the Program

Figure 3-1 provides an overview of how the evaluated projects were assigned their level of engineering analysis rigor. The level of rigor assigned to each project was based on the evaluation manager's review and consideration of the complexity of the analysis, the availability of or ability to obtain data, the magnitude of savings, customer sensitivities, and overall budget available.

Figure 3-1. Method for Assignment of Rigor

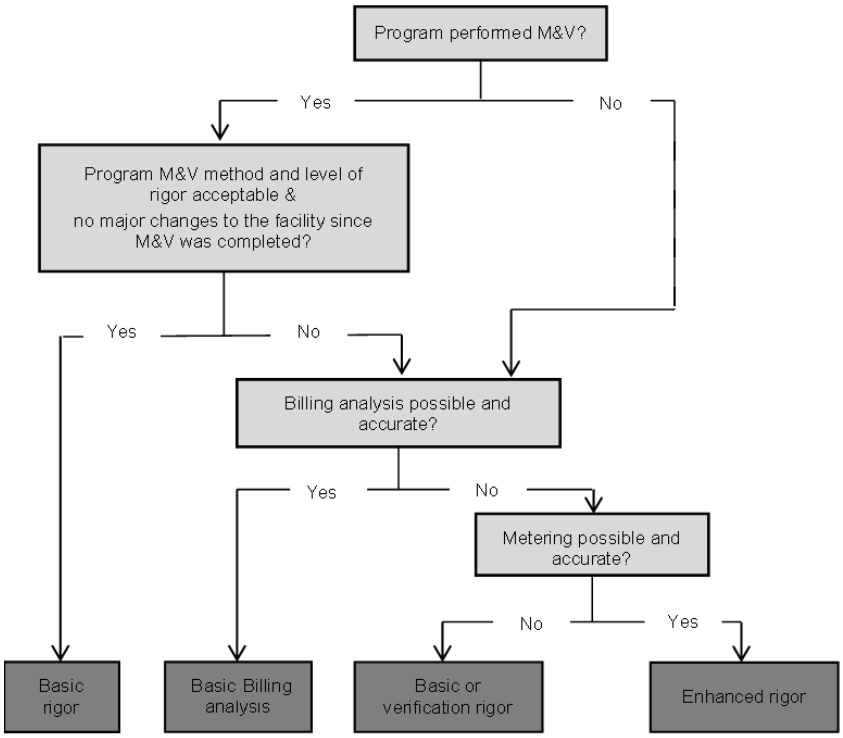


Table 3-1 summarizes the levels of rigor employed in the evaluation.

Table 3-1. Level of Rigor for Sampled Projects

Level of Rigor	Description	Planned Percentage of Projects	Actual Percentage of Projects
Verification	Inspection or review-only verification. If pertinent equipment is not accessible or cannot be metered over the long term, the Impact Evaluation Team uses a combination of research-based methods to determine evaluated gross savings. The Impact Evaluation Team verifies the installation of all or a sample of project measures through on-site inspection. To calculate savings, the Impact Evaluation Team relies on deemed savings references, measure evaluation results from other similar projects in the sample, and/or past equipment-specific performance studies either conducted or supported by NYSERDA. Verification is for those projects where a basic or enhanced level of rigor is not possible.	10%	2%
Basic	Typically consistent with IPMVP Option A (partially measured retrofit isolation)[1]. Analysts design M&V plans around spot measurements and short-term metering to supplement previously performed Program M&V with additional data collection as needed to enhance the previously reported results to the level of accuracy required for evaluation M&V. Option C is preferred where it is feasible.	20%	26%
	Typically consistent with IPMVP Option C (whole-facility level analysis); an analysis based on weather-normalized billing analysis often is this level of rigor. If projected savings exceed about 10% of the pre-installation utility bills for a specific project, bill data is available, and savings can be normalized for production, then billing analysis is used to assess the change in use for that project. Utility meters or customer submeters that are dedicated to the specific buildings being evaluated are used if they can better isolate the energy use.		31%
Enhanced	Typically consistent with IPMVP Option B (retrofit isolation) or Option D (calibrated simulation level analysis) and also including time-of-use light logger-based M&V; this level of rigor usually includes metering of end-use energy for between 4 and 8 weeks with advanced modeling of the building or process, calibrated against field measurement of specific equipment. Enhanced rigor projects may also include whole-building simulations for projects where such methods are appropriate, such as comprehensive new construction efforts.	70%	40%

[1] Refer to Table 2: Overview of M&V Options in International Performance Measurement & Verification Protocol, Concepts and Options for Determining Energy and Water Savings, Volume I, January 2012.

The only projects evaluated at a verification level of rigor were those for which higher rigor was not possible and data was not available to do the other analysis options. Based on the actual measure mix for the sampled sites, the level of rigor changed. Billing analysis turned out to be the most appropriate means of analysis for almost a third of the projects, especially those with natural gas measures. This increased the final proportion of basic rigor projects compared to the plan.

Also, site energy management system data was more available than projected. Expanded use of such data (and corresponding commitment to be less invasive conducting evaluation M&V at those sites) also increased the proportion of basic rigor projects over plan.

3.2 SAMPLE DESIGN

The sampling started very early, before the development of the formal work plan, due to weather-related constraints on the overall evaluation time frame. A sampling memo for natural gas saving projects, with savings occurring predominantly in the winter, was presented to DPS for approval at the start of the evaluation project. This sampling memo is provided as Appendix E of this document. The impact team also received a separate approval from DPS to start summer metering work in advance of the development of a formal work plan, also for seasonal reasons.

The evaluation scope changed after an original sample design was completed, the sample was pulled, and project files were in hand. The electric Program scope was revised to evaluate custom projects only, which eliminated 39 prequalified electric sites. As a result, the electric on-site M&V sample size decreased from 100 sites to 61 sites. The revised scope also excluded attribution research.

3.1.1 Precision and Bias

For natural gas projects the targeted relative precision was 6% on evaluated gross savings. For custom electric projects 10% relative precision was targeted.

3.2 ON-SITE M&V SAMPLE DESIGN

Stratified ratio estimation (SRE) was used for the sample design because it allows for efficient sampling design and generally requires a lower sample size for a targeted level of precision as long as there is an expected strong correlation between the program-reported savings and the evaluated gross savings. A summary of the sampling plan is represented in Table 3-2 and followed by a discussion of select components.

Table 3-2. Summary of the Sampling Plan

Sampling Component	Sample Approach	Comments
Two sample frames: one for the electric population, one for the natural gas population	Program-reported data, all projects with at least one measure installed between January 1, 2012, and September 30, 2013	Program-reported data was provided by NYSERDA.[1]
Method	Stratified ratio estimation	Correlation between program-reported and evaluation savings is expected to be strong.
Error ratio (ER)	ER of 0.50 for electric projects ER of 0.60 for natural gas projects	The kWh ER from the previous (2009–2010) EF evaluation. The MMBtu ER project was derived from natural gas impact evaluations in another jurisdiction.
Variable to estimate	RR for annual electric (kWh) or natural gas (MMBtu) savings	M&V will establish evaluated savings. RR is calculated as the ratio of the evaluated savings to the program-reported savings.
Primary sampling unit	Smallest combination of project and site	Two projects at a site constitute two sampling units, as does one project with measures installed at two different sites. This approach was adopted to enable efficient implementation of the on-site M&V.
Upper level stratification variables	Program administrator (NYSERDA, NFGDC). NYSERDA was stratified by fuel type (electric, gas), and Upstate vs. Downstate	Separate sampling for NYSERDA and National Fuel. NYSERDA was then stratified by fuel type; The NYSERDA program was further stratified into Upstate and Downstate.
Lower level stratification variables	Size	Size determined by the annual kWh savings (for project sites with electric savings) and MMBtu savings (for project sites with natural gas savings).

[1]The initially drawn electric sample frame included both custom and prequalifed measures, however subsequently, the prequalified sites were dropped.

3.2.1 Primary Sampling Unit

The sampling unit is the smallest combination of project and site. If a single site hosts two projects, each project is a different sampling unit, and if a single project application covers installation at two different sites, then each site is a different sampling unit. The program tracking data is available for Impact Evaluation Team to systematically and uniformly make this separation. “Site” is defined as a single location that may include one or multiple buildings in close proximity (e.g., a premise). This approach was adopted to enable efficient implementation of the on-site M&V work.

3.2.2 Stratification

The initial stratification was by Program administrator (i.e., NYSERDA and NFGDC). The sample frames were stratified in order to facilitate production of RR for particular participant groups (upper level stratification) and to improve sampling efficiency (lower level stratification).

Upper Level Stratification

The upper stratification variables for the electric and natural gas on-site sample are described below.

Electric projects:

- Downstate electric – All of the projects in Con Edison territory
- Upstate electric – All projects not in Con Edison territory

Natural gas projects:

- NFGDC natural gas (serves upstate exclusively). The NFGDC program is administered by NYSERDA, but the savings are reported by NFGDC
- NYSERDA Upstate natural gas (excludes NFGDC)
- NYSERDA Downstate natural gas

Each project was assigned to a single category, based on the project or measure type shown in NYSERDA's tracking database.

Lower-Level Stratification

Size is used as the lower-level stratification variable. Size categories were based on the magnitude of the reported annual savings. Cutoffs were established using the method described in the 2004 *California Evaluation Framework*.[5]

Projects in the smallest size stratum, which in all cases included less than 3% of the total energy savings for the upper-level stratification category, were not evaluated. While there are many of these small projects, they account for a small part of the overall program impacts and will have little effect on the RR. The RR developed for the rest of the sample frame was applied to these projects.

[5] TecMarket Works, et. al. *The California Evaluation Framework*. Project Number: K2033910. Prepared for the California Public Utilities Commission and the Project Advisory Group. June 2004. Pages 327 – 339 and 361 – 384.

3.2.3 Final Sample Disposition

Table 3-3 presents the custom electric participant completed sample, broken out by upper- and lower-level stratification variables.

Table 3-3. NYSERDA Electric Projects – Upper- and Lower-Level Stratification Results[1]

Upper-Level Stratum	Lower-Level Stratum	# of Sites	Mean kWh Savings	Sample Size	Final Sample Disposition	Sample Percentage of Total Sites	Target Relative Precision at 90% Confidence Level
Upstate	1	37	1,029,740	12	13	32%	14%
	2	65	280,476	10	11	15%	
	3	68	86,065	3	3	4%	
	4	183	23,659	6	5	3%	
	5[a]	*49*	*7,264*	*0*	*NA*	*0%*	
	Total[a]/ Mean[a]	**353**	**188,423**	**31**	**32**	**9%**	
Downstate	1	13	1,026,339	7	7	54%	13%
	2	26	444,965	10	9	38%	
	3	47	181,676	10	10	21%	
	4	55	47,803	3	4	5%	
	5[a]	*32*	*9,897*	*0*	*NA*	*0%*	
	Total[a]/ Mean[a]	**141**	**255,883**	**30**	**30**	**21%**	
Statewide Custom Electric		**494**	**207,677**	**61**	**62**	**12%**	**10%**

[1] Of the electric saving sampled projects, one includes program-reported fuel savings.

[a] Projects in the smallest size stratum (5), which in all cases included less than 3% of the total energy savings for the upper-level stratification category, were not evaluated and was not included in the overall mean calculation.

Table 3-4 provides more information on the disposition of custom electric sites. Overall, the completion response rate was 74%, and the cooperation rate was 90%,

Table 3-4. NYSERDA Electric Projects –Sample Disposition

Disposition	# of Sites
Cooperated complete	62
Partial interview, incomplete	4
Dropped - Site Complications	2
Dropped - Target Reached	4
Refusal	8
Non-contact/unresponsive	4
Total	**84**

Table 3-5 presents the natural gas participant samples, broken out by upper- and lower-level stratification variables. The final sample disposition is noted as well.

Table 3-5. Natural Gas Projects – Upper- and Lower-Level Stratification Results[1, 2]

Upper-Level Stratum	Lower-Level Stratum	# of Sites	Mean MMBtu Savings	Sample Size	Final Sample Disposition	Sample Percentage of Total Sites	Target Relative Precision at 90% Confidence Level
NFGDC	1	7	15,981	7	7	100%	**11%**
	2	24	10,185	12	12	50%	
	3	98	5,987	12	12	12%	
	4[a]	*78*	*985*	*0*	*NA*	*0%*	
	Total[a]/ Mean[a]	**129**	**7,310**	**31**	**31**	**24%**	
Upstate	1	3	24,528	3	3	100%	**11%**
	2	7	14,721	6	6	86%	
	3	16	10,926	6	7	38%	
	4[a]	*28*	*1,537*	*0*	*NA*	*0%*	
	Total[a]/ Mean[a]	**26**	**13,517**	**15**	**16**	**58%**	
Downstate	1	2	19,560	2	2	100%	**10%**
	2	2	3,790	2	2	100%	
	3	5	6,024	3	5	60%	
	4[a]	*5*	*362*	*0*	*NA*	*0%*	
	Total[a]/ Mean[a]	**9**	**8,536**	**7**	**9**	**78%**	
Statewide Natural Gas Mean/Total		**164**	**8,362**	**53**	**56**	**32%**	**6%**

[1]Of the gas savings sampled projects, nine include program-reported electric savings.

[2]There are four projects that are in both gas and electric samples.

[a] Projects in the smallest size stratum (4), which in all cases included less than 3% of the total energy savings for the upper-level stratification category, were not evaluated and was not included in the overall mean calculation.

Table 3-6 provides more information on the disposition of the prequalified and custom natural gas sites. Overall the completion response rate was 67%, and the cooperation rate was 81%.

Table 3-6. Natural Gas Projects – Sample Disposition

Disposition	# of Sites
Cooperated complete	56
Partial interview, incomplete	2
Dropped – site complications	1
Dropped – target reached	2
Refusal	7
Non-contact/unresponsive	5
Total	**73**

3.3 METHOD FOR REALIZATION RATE DEVELOPMENT

The objective of this impact evaluation is to estimate the evaluated electric and natural gas savings for Program projects installed in the period from 2010–2013; primary research for

program years 2012-2013 was conducted, but a calculated RR was developed for years 2010-2011 using prior and current evaluation results. The electric program was evaluated in the prior impact evaluation for the period from 2006–2009. Since the Program remained relatively stable from 2006 through 2014, the primary data collection targeted projects installed in the last two years of the period (specifically, January 1, 2012 – September 30, 2013). This approach keeps the evaluation as relevant as possible by focusing on more recent implementation at the same time that it reduces study costs. In addition, the natural gas programs did not begin until 2012; this approach keeps the time frames consistent for the electric and natural gas sample frames.

The Impact Evaluation Team computed the aggregated weighted electric energy and demand RRs for prequalified projects from the 2006–2009 impact evaluation data (they had not previously been reported separately from custom projects) and used these two values to deem RRs for subsequent periods.

Table 3-7 summarizes that basis for the RRs to assign to projects completed for the various periods.

Table 3-7. EFP Source of Primary Data Collection for Realization Rates

	Basis of Realization Rate Estimate for Projects Installed during the Period		
EFP Program	**1/12006–12/31/2009**	**1/1/2010–12/31/2011**	**1/1/2012–9/30/2013**
Electric prequalified	Prior evaluation	Prior evaluation	Prior evaluation
Electric custom	Prior evaluation	Blended RR	Current evaluation
Natural gas prequalified & custom	*Program did not exist*	*Program did not exist*	Current evaluation

SECTION 4: RESULTS, FINDINGS, AND RECOMMENDATIONS

The section presents the results and findings from the evaluation. The section concludes with recommendations.

4.1 ELECTRIC ENERGY SAVINGS RESULTS

This section summarizes the results of the measurement and verification (M&V) activities for electric energy projects from January 1, 2012, through September 30, 2013.

4.1.1 Program Electrical Energy Savings and Realization Rates for Custom Projects

The RR of 1.01 for the electric custom projects installed in the period of January 1, 2012, through September 30, 2013, is calculated by dividing the evaluated gross savings by the program-reported savings. Table 4-1 provides a comparison of the program-reported and evaluated electric energy savings for custom projects.

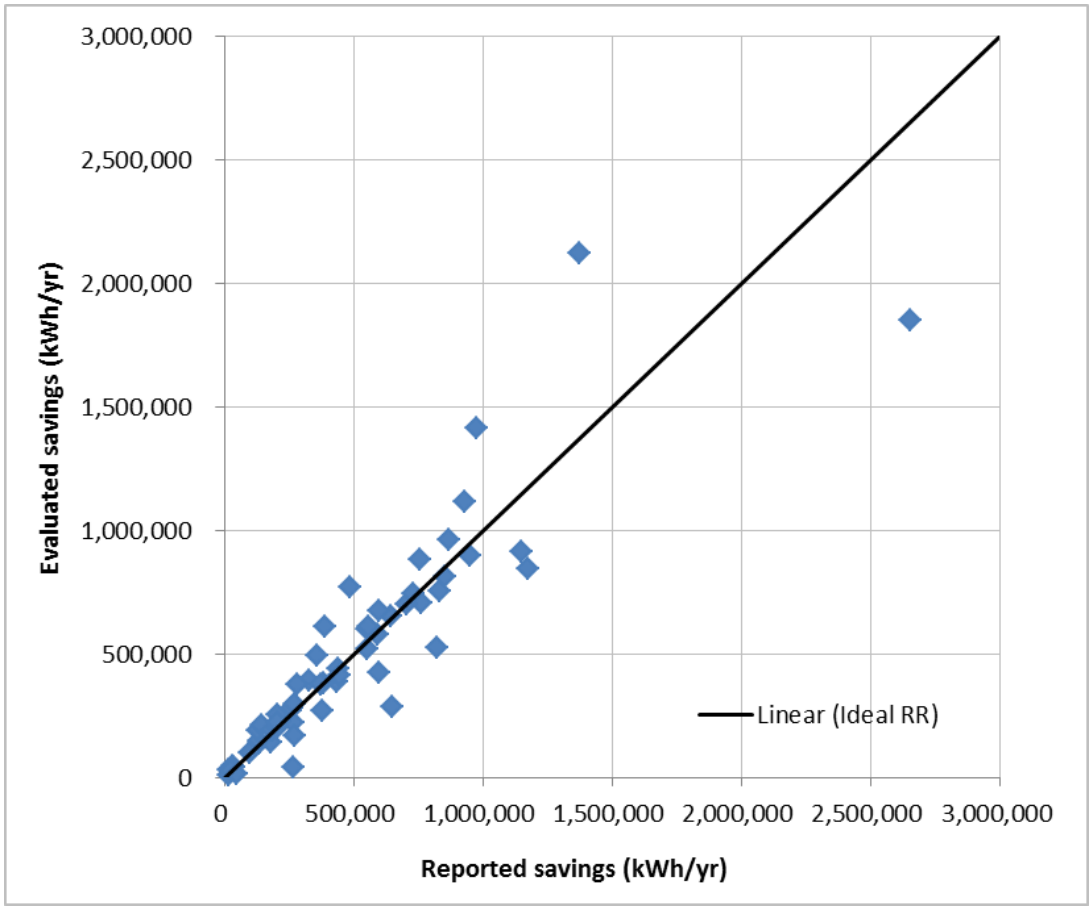
Table 4-1. NYSERDA Reported and Evaluated Electric Energy Savings (Custom projects)

Parameter	Program-Reported Savings	Realization Rate	Evaluated Gross Savings[1]	Relative Precision
Electric energy (MWh/yr)	103,266	1.01	104,299	6.4%
Electric demand (MW)	17.1	0.98	16.8	13.2%

[1] Realization rate and evaluated savings are rounded.

In Figure 4-1 below, the ideal result where the evaluated gross savings would always match the program-reported savings, i.e., a RR of 1, is shown as a solid black line. Actual findings are plotted as points on the graphs. A pattern of points below the ideal line illustrates an RR of less than 1; points above the line illustrate an RR greater than 1.

Figure 4-1. NYSERDA Program-Reported and Evaluated Custom Electric Energy Savings



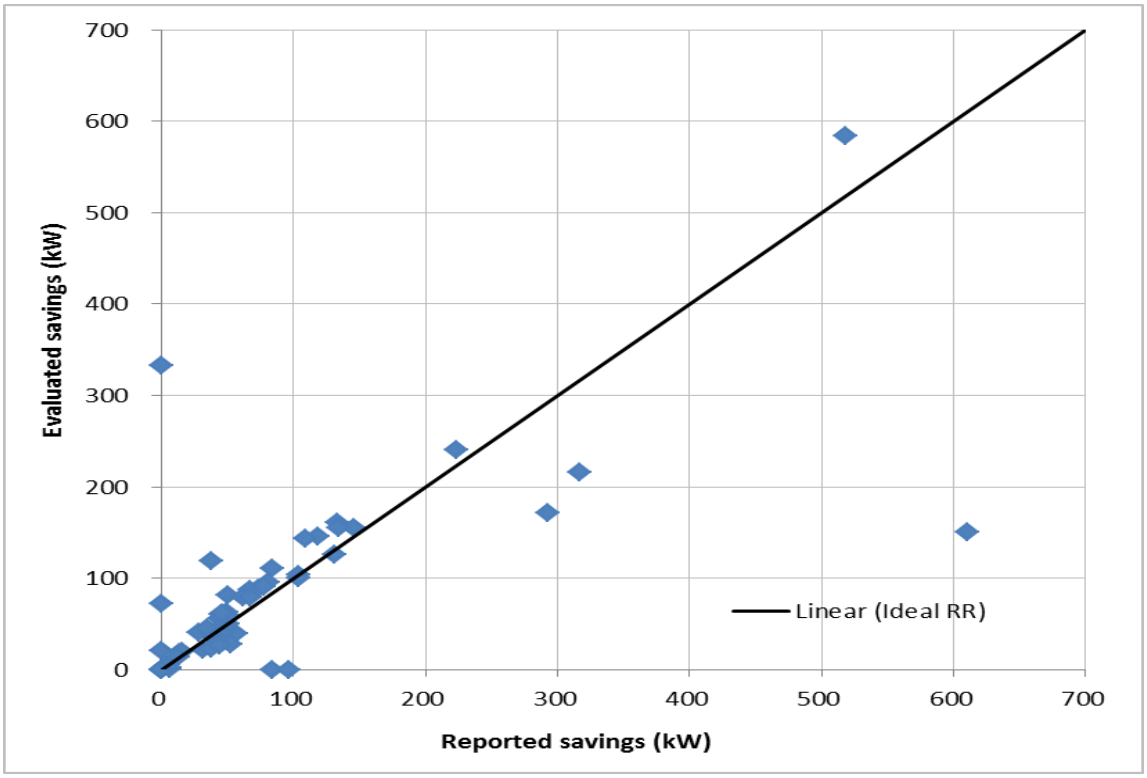
The error ratio is the key measure of the variability in the relationship between the evaluated estimate of gross savings[6] and the program reported savings. The larger the error ratio, the larger the sample size required to meet the targeted level of confidence and precision. Conversely, the smaller the error ratio, the smaller the sample size required to meet the targeted level of confidence and precision[7]. The electric energy efficiency RR error ratio is 0.31. This is a relatively low value. In particular it is lower than the sample design assumed, which led to evaluation results with better than planned level of confidence and precision.

[6] See the Glossary for additional technical detail.

[7] TecMarket Works. (2004). *The California Evaluation Framework*. Prepared for the California Public Utilities Commission and the Project Advisory Group, p. 332.

Figure 4-2 illustrates the evaluated annual electric coincident peak demand savings compared to that reported by the Program. As described above, the ideal is shown as a solid black line on the charts. A pattern of points below the ideal line illustrates an RR of less than 1; points above the line illustrate an RR greater than 1.[8]

Figure 4-2. NYSERDA Program-Reported and Evaluated Custom Coincident Peak Demand Savings



Upper-Level Stratification Category Results

The evaluation sample included two upper-level stratification categories: upstate and downstate. These categories were chosen through discussions with Program staff. As shown previously in Table 3-3, the custom upstate and downstate categories were expected to have target precision values of 0.14 and 0.13, respectively. Table 4-2 provides the evaluation results by upper-level stratification category.[9]

[8] Although the program-level custom electric coincident peak demand savings RR is 1.02, the dotted line in the plot suggests an RR lower than 1. This is because the dotted line on the plot is a trend line and the case weights are not taken into account.

[9] For this and similar detailed analysis the evaluation focused on electric energy savings over demand savings.

Table 4-2. NYSERDA Electrical Energy Results by Upper-Level Stratification (Custom projects)

Description	Program-Reported Savings (MWh/yr)	Realization Rate	Evaluated Gross Savings[1,2] (MWh/yr)	Relative Precision
Downstate	36,397	1.04	37,853	8.5%
Upstate	66,869	1.00	66,869	5.3%
Overall	**103,266**	**1.01**	**104,299**	**6.4%**

[1] Realization rate and evaluated savings are rounded.
[2] Due to rounding the totals do not match.

There was no clear pattern to the RR with respect to geographic region or size.

4.1.2 Differences between Program-Reported and Evaluated Electrical Energy Savings

For each project with an evaluated RR other than 1, a difference analysis was performed to identify the major driver or drivers to the evaluated RR. The results from the difference analysis are aggregated in an attempt to identify systematic differences. The differences were divided into primary categories and subcategories within the primary categories. The primary category indicated the phase of the project when the difference in savings was likely to occur. The subcategories within the primary categories further provided detailed reason for the difference in savings.

The results of the difference analysis are presented in Table 4-3. This table shows the positive and negative savings impact. Differences that increased RR are shown as the blue bars, and differences that reduced RR are shown as the red bars.

Table 4-3. NYSERDA Differences: Analysis Results – Electric Energy (Custom projects)

Discrepancy Category	Discrepancy Sub-Category	Counts	Positive and Negative Impact on RR*	Overall Impact on RR*	Total kWh Impact
Application review	Difference in as-built equipment efficiency	1			14,373
	Difference in cooling or heating interactivity	9			2,587,320
	Difference in equipment hours of operation	7			-102,256
	Inaccurate estimation from applicant model	13			-469,945
	Inaccurate normalization to typical weather	4			-259,023
	Inaccurate pre-project characterization	5			-2,813,851
	Incorrect baseline reference	2			-3,119,460
	Ineligible measure	1			-341,087
	Insufficient assessment of measure interactivity	5			1,568,570
	Unknown applicant algorithm or assumptions	1			-2,996
Measure installation verification	Difference in cooling or heating interactivity	2			108,203
	Difference in installed control strategy	2			132,060
	Difference in installed equipment size	3			-55,381
	Difference in installed equipment technology	2			384,105
	Difference in quantity installed	9			836,433
Measure performance	Difference in cooling or heating interactivity	18			5,142,293
	Difference in equipment hours of operation	57			-444,212
	Difference in equipment load profile	11			-779,258
	Difference in installed equipment efficiency	13			-215,485
Tracking	Tracking database inconsistency	11			-1,288,647
Total reported savings = 103,266 kWh					

* The blue bars indicate positive savings whereas the red bars indicate negative savings.

Table 4-3 above demonstrates key contributors to the 1.01 RR for electric energy savings. Note that the overall positive and negative savings impacts of the differences negate each other.

Here is a discussion of some of the major drivers to RR:

- Different operating characteristics (measure performance) were the largest difference category driving up the RR. This category includes differences in installed equipment efficiency, load profile, and run hours, and equipment interactivity.
    - Cooling interactivity was responsible for the majority of the positive influence on electric RR.
    - The individual positive and negative savings impacts due to a difference in operating hours were large. However, they canceled each other out, resulting in negligible influence on the electric RR.
- Measure installation verification is the second-largest category driving up the RR. This category includes differences in installed technologies, quantities, or control strategies.
    - The primary driver in this category was found to be the change in installed measure quantities for nine measures.

- Application review is the largest difference category driving down the RR. This category includes differences in baseline efficiencies, pre-project characterization, and installation of ineligible measures.
    - Two projects with differences in baseline efficiency accounted for 44% of the overall negative impact in this category.
- These differences generally represent stand-alone events. The analysis found no systematic differences when it was compared with the evaluation.

4.1.3 Additional Electrical Energy Savings Results

The Impact Evaluation Team also reviewed the evaluated projects to determine if patterns could be identified or feedback could be provided about project performance based on measure type, building type, and geographic location. Although these results do not adhere to the same 90/10 confidence precision targets as the evaluation sample, the results are suggestive regarding program areas of strength and opportunities for refinement regarding measure savings estimation.

Custom Electric Energy Savings by Measure and Building Type

The variable that most affected the electric energy RR was measure type. For example, space cooling equipment and related controls averaged a 31% RR, 70% less than any other electric measure type. Conversely, lighting RRs were consistently above average, with the preponderance of the offsetting cooling measure underperformance. The underlying cause of the understated lighting measure savings was the absence of accounting for the added benefit of reduced air conditioning load due to reduced lighting heat gain in applications. Based on discussions with the program staff, it is not cost-effective to include interactive savings for small lighting measures. This particular issue was also discussed with the program staff at length at the conclusion of the previous EFP evaluation and the consensus was that there is not easy solution to address this issue. As a result no changes were made to the application process to incorporate the interactive savings.

Regarding building type, the RRs for projects in commercial facilities tended to exceed those of non-profit facilities. Otherwise, building type did not reveal electric RR laggards or leaders after adjusting for measure type.

Appendix B includes a detailed analysis of realization rates by measure and building type.

Custom Electric Energy Savings for Update and Downstate Projects

NYSERDA has historically been interested in whether there are material differences in the performance of projects located in upstate New York and downstate New York. The Impact Evaluation Team classified each of the projects in the sample frame into one of these two categories based on their electrical utility[10] and determined that 65% of the program-reported custom project savings was attributable to upstate projects while 35% of the program-reported custom project savings was attributable to downstate projects. The Impact Evaluation Team then reviewed the electrical energy savings of the sampled sites and calculated the upstate project RR to be 99% and the downstate project RR to be 101%. The reason for the small difference in RRs is driven by the higher number of cooling-related projects in the upstate region. This category of project had lower RRs than either the lighting or VFD projects, resulting in a lower apparent upstate RR. Based on this, and previous review of the impact evaluation results by project type, the Impact Evaluation Team does not believe there is a correlation between project location and RR for the EFP program.

4.1.4 Program Electrical Energy Savings and Realization Rates for Prequalified Projects

As described in the methodology, the Impact Evaluation Team applied the 2006–2009 prequalified electric energy and coincident peak demand RRs to the 1/1/2012–9/30/2013 period's reported savings to calculate the evaluated gross savings. The results are summarized in Table 4-4.

Table 4-4. Reported and Evaluated Electric Energy Savings (Prequalified projects)

Parameter	Program-Reported Savings	Realization Rate	Evaluated Gross Savings	Relative Precision
Electric energy (MWh/yr)	118,207	1.00	118,207	N/A
Electric demand (kW)	14.4	1.02	14.7	N/A

4.2 NATURAL GAS SAVINGS RESULTS FOR NYSERDA

This section summarizes the results of the M&V activities for natural gas projects. In 2012 funding for natural gas energy savings projects became available; as such, natural gas project participation has increased sufficiently since the last evaluation cycle to warrant evaluation.

[10] The Impact Evaluation Team classified all projects served by Con Edison as downstate projects. All other projects were classified as upstate projects.

4.2.1 Program Natural Gas Savings and Realization Rates

The RR for the 2012–2013 natural gas energy savings of the Program, calculated as the evaluated gross savings divided by the program-reported savings, is 0.64. Table 4-5 provides a comparison of the reported and evaluated natural gas energy savings.

Table 4-5. NYSERDA Reported and Evaluated Natural Gas Energy Savings

Parameter	Program-Reported Savings	Realization Rate[1]	Evaluated Gross Savings	Relative Precision
Natural gas (MMBtu/yr)	81,448	0.64	52,127	7.9%

[1] Due to program changes that occurred in 2014, the Impact Evaluation Team recommends that NYSERDA use an RR of 0.63 as a prospective pro forma value for 2014 and later reporting, until the next evaluation is completed.

Figure 4-3 illustrates the evaluated annual natural gas energy savings compared with those reported by the Program. Ideally, the evaluated gross savings would always match the program-reported savings. This ideal is shown as a solid black line on the charts. The actual findings are plotted as points on the graphs. Figure 4-4 presents the same information on a magnified scale.

Figure 4-3. NYSERDA Program-Reported and Evaluated Natural Gas Energy Savings

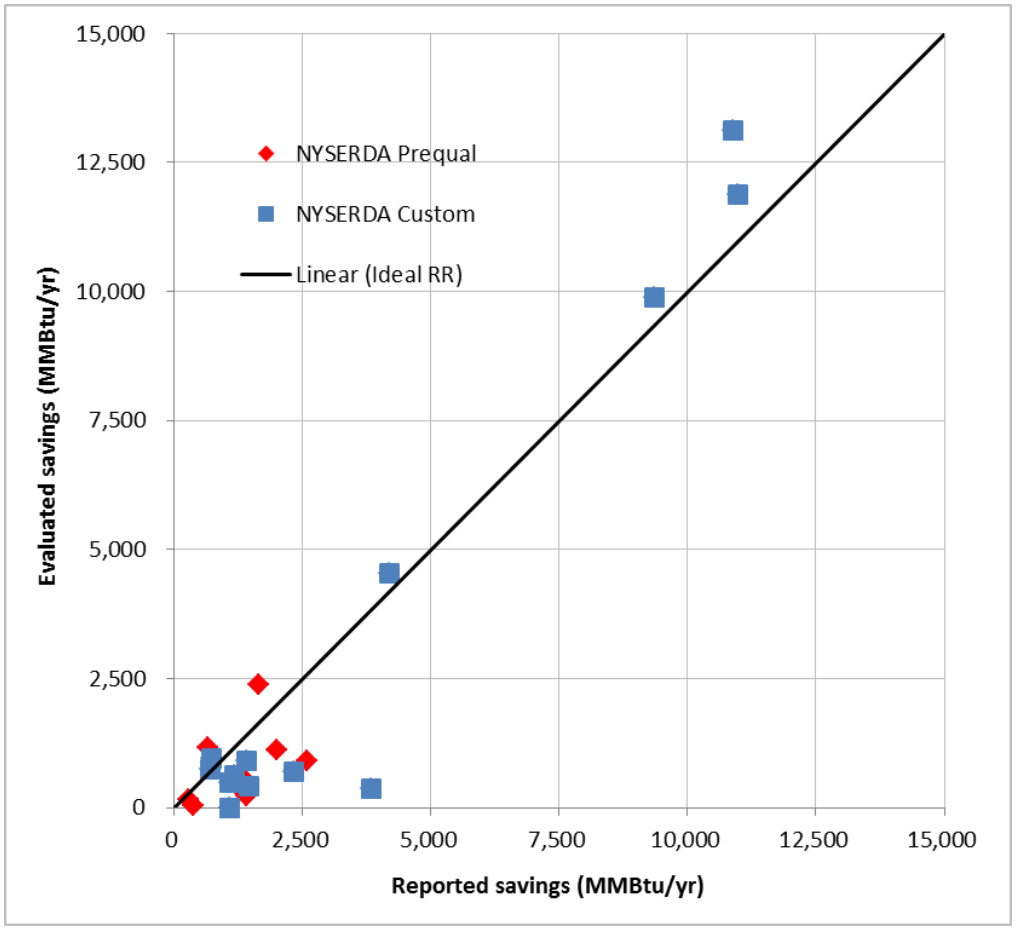
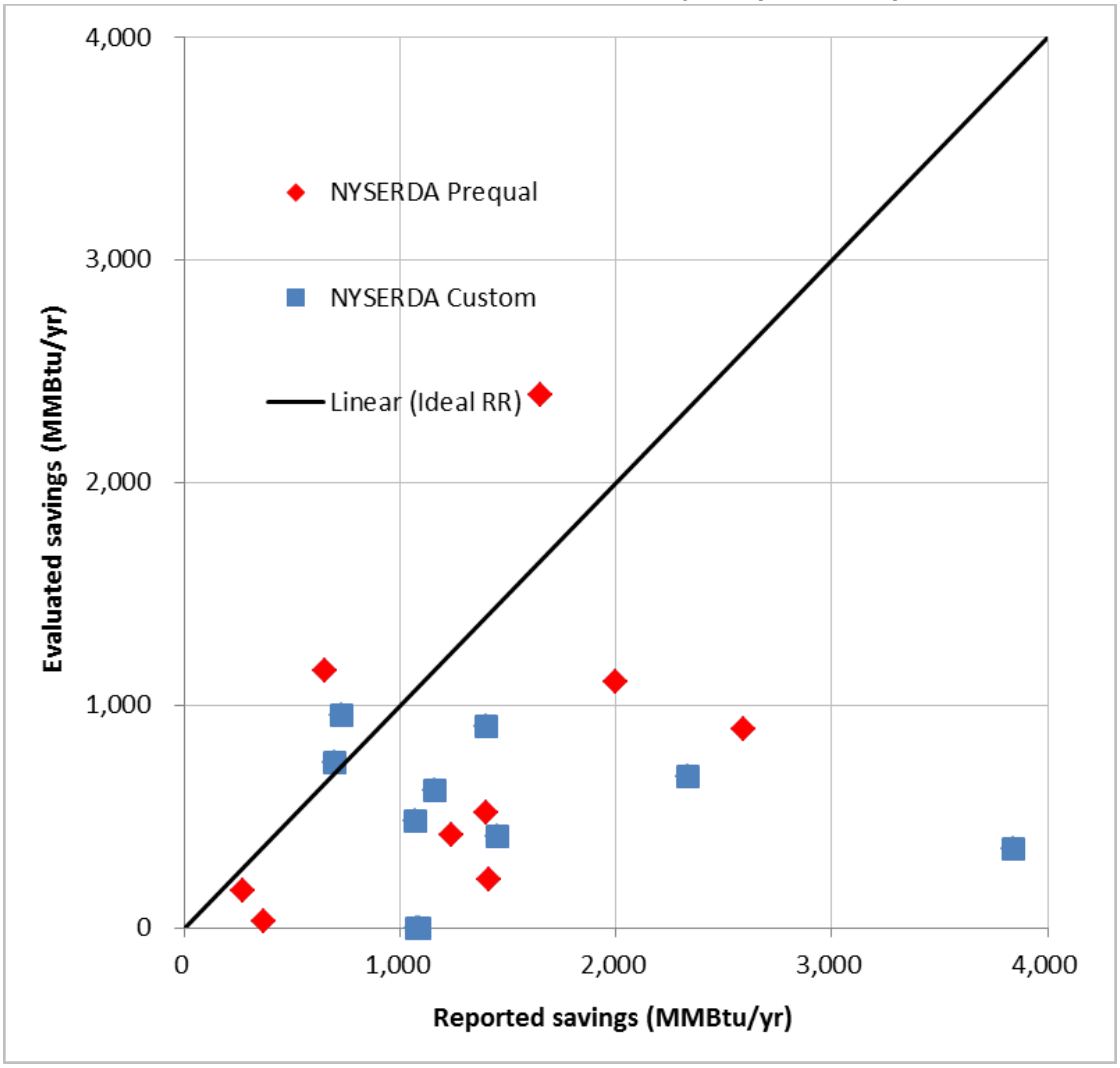


Figure 4-4. NYSERDA Program-Reported and Evaluated Natural Gas Energy Savings, *Reduced Axis Scale for Detailed View of Smaller Projects* (NYSERDA)



About 80% of NYSERDA's reported natural gas savings are from custom projects. As the full scale graph illustrates, the four largest projects, all custom, had good RRs, close to and slightly in excess of 1. The evaluated savings for smaller custom projects and the prequalified projects varied a lot compared to the reported savings. This led to a high natural gas RR error ratio of 1.02.

Upper-Level Stratification Category Results

The evaluation sample included two upper-level stratification categories: upstate and downstate. These categories were chosen through discussions with Program staff. As shown previously in Table 3-4, the upstate and downstate categories were expected to have a target precision value of 11% and 10%, respectively, after excluding the prequalified projects. Table 4-6 provides the evaluation results by upper-level stratification category.

Table 4-6. NYSERDA Natural Gas Results by Upper-Level Stratification

Description	Program-Reported Savings (MMBtu/yr)	Realization Rate	Evaluated Gross Savings[1,2] (MMBtu/yr)	Relative Precision	
Downstate	29,736	0.34	10,110	0%	
Upstate	51,712	0.80	41,887	10.9%	
Program-level	**81,448**	**0.64[3]**	**52,127**	**7.9%**	

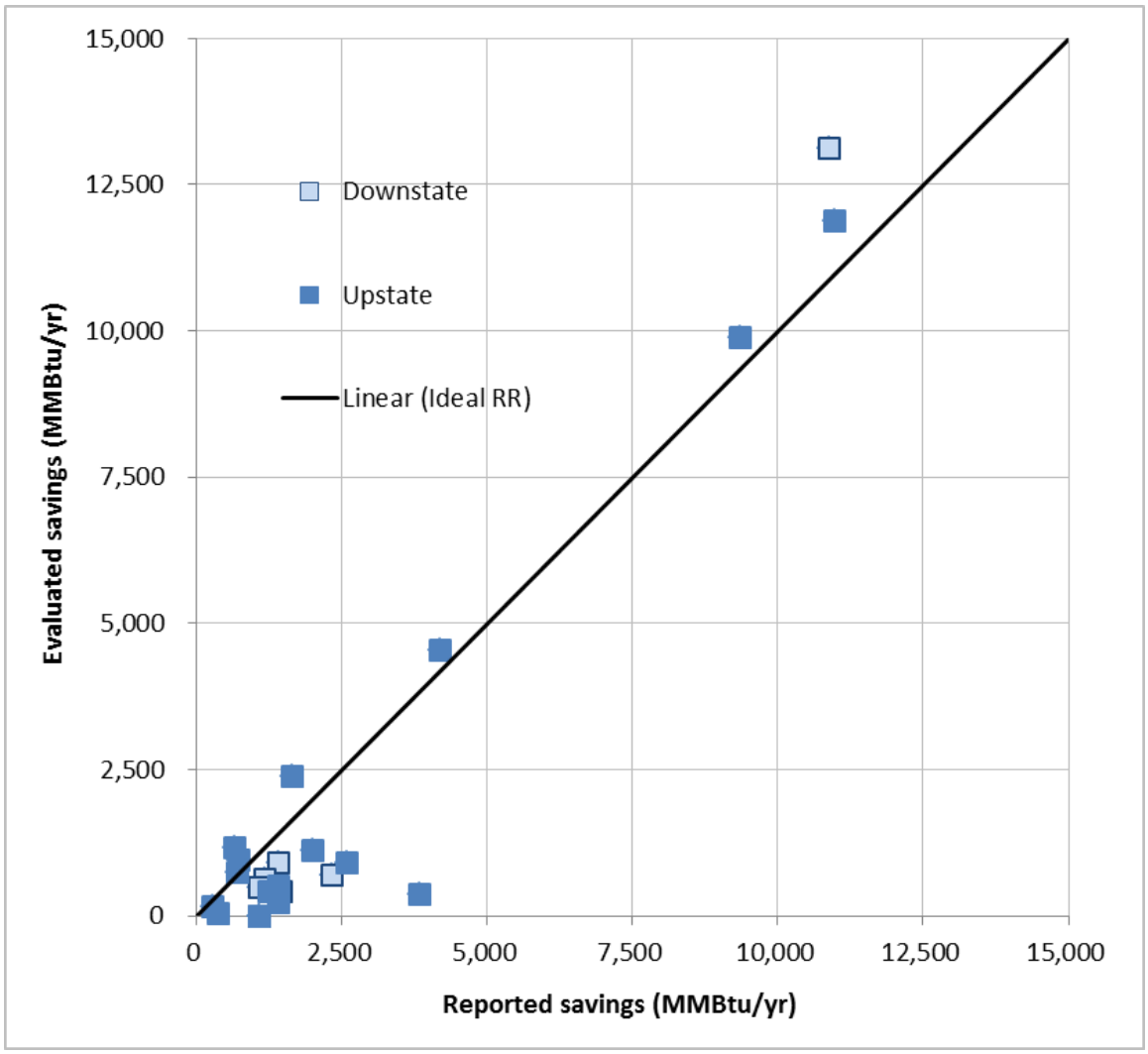
[1] Realization rate and evaluated savings are rounded.

[2] Due to rounding the totals do not match.

[3] Due to program changes that occurred in 2014, the Impact Evaluation Team recommends that NYSERDA use an RR of 0.63 as a prospective pro forma value for 2014 and later reporting, until the next evaluation is completed.

Downstate sites constituted 36% of the sample, and were all small custom projects. Conversely, the upstate sites had a good mix of custom and prequalified sites. This Upstate group had all the large custom sites with RRs close to 1, which contributed to a better overall RR for the upstate strata. Overall, the large savings projects had better RRs compared to the small savings projects. The large savings projects tend to be custom projects which include site specific analysis. The small savings projects typically tend to be prequalified projects which means they use deemed savings numbers. Deemed savings numbers are not site specific and hence result in savings discrepancies when evaluated. Figure 4-5 illustrates the evaluated savings for upstate and downstate sites.

Figure 4-5. Reported and Evaluated Natural Gas Energy Savings by Region (NYSERDA)



4.2.2 Differences between Program-Reported and Evaluated Natural Gas Savings

For each project with an evaluated RR other than 1, a difference analysis was performed to identify the major driver or drivers to the evaluated RR. As with the electric projects, the difference analysis results are aggregated in an attempt to identify systematic differences in approach between the Program and the Impact Evaluation Team. The results of the difference analysis are presented in Table 4-7.

Table 4-7. Differences Analysis Results – Natural Gas (NYSERDA)

Discrepancy Category	Discrepancy Sub-Category	Counts	Positive and Negative Impact on RR*	Overall Impact on RR*	Total MMBtu Impact
Application review	Difference in as-built equipment efficiency	2			-407
	Difference in equipment hours of operation	1			-686
	Inaccurate estimation from applicant model	12			-10,162
	Inaccurate normalization to typical weather	2			1,246
	Inaccurate pre-project characterization	6			-15,910
	Incorrect baseline reference	3			-1,358
	Insufficient assessment of measure interactivity	1			-58
	Unknown applicant algorithm or assumptions	1			-777
Measure installation	Difference in installed control strategy	1			-103
	Difference in quantity installed	2			488
Measure performance	Difference in equipment hours of operation	12			1,004
	Difference in equipment load profile	7			-6,626
	Difference in installed equipment efficiency	10			3,326
Total reported savings = 81,448 MMBtu					

* The blue bars indicate positive savings whereas the red bars indicate negative savings.

Table 4-7 above also illustrates the key contributors to the 0.64 RR for natural gas energy savings. Some of the major drivers to RR are discussed below:

- The application review category has the largest impact on RR. This category includes differences in baseline efficiencies, pre-project characterization, and installation of ineligible measures.
    - Differences in pre-project characterization such as difference in flow rate or pre-installation energy usage resulted in the bulk of the differences associated with this category. For example, for one site the Impact Evaluation Team found that the overall pre-installation usage was greater than that estimated by the applicant bin-analysis model.
    - In addition, a systematic error in the furnace boiler prequalified savings calculation tool was identified that resulted in reduced a RR for every project that depended on it. The tool used the input boiler capacity where it should have used output capacity, resulting in overestimation of savings. However, this issue was fixed by Program staff in 2014 and all the projects after the tool fix used the correct algorithm. Adjusted RRs were calculated for use going forward.
- Measure performance category is the second-largest difference category driving down the RR. This category includes differences in installed equipment efficiency, equipment interactivity, load profile, and run hours.

- Within this category, differences in equipment load influenced the natural gas RR the most. For example, at a large scale laundry facility, the daily output of the washer was not quite as large as anticipated by the applicant resulting in lower average loads per day.
- The installed equipment efficiency was observed to positively affect the natural gas RR. For example, for one site the applicant used an installed boiler efficiency of 98% while the testing and calibration documents provided by the facility contact showed the boiler to operating at an efficiency of 91%.

4.2.3 Additional Natural Gas Savings Results

Appendix C includes a detailed analysis of NYSERDA natural gas realization rates by measure and the combination of NYSERDA and NFGDC results by building type.

4.3 NATURAL GAS SAVINGS RESULTS FOR NFGDC

This section summarizes the results of the M&V activities for NFGDC natural gas projects.

4.3.1 Program Natural Gas Savings and Realization Rates

The RR for the NFGDC is 0.66; the calculation is the evaluated gross savings divided by the program-reported savings. Table 4-8 provides a comparison of the NFGDC's program-reported and evaluated natural gas energy savings.

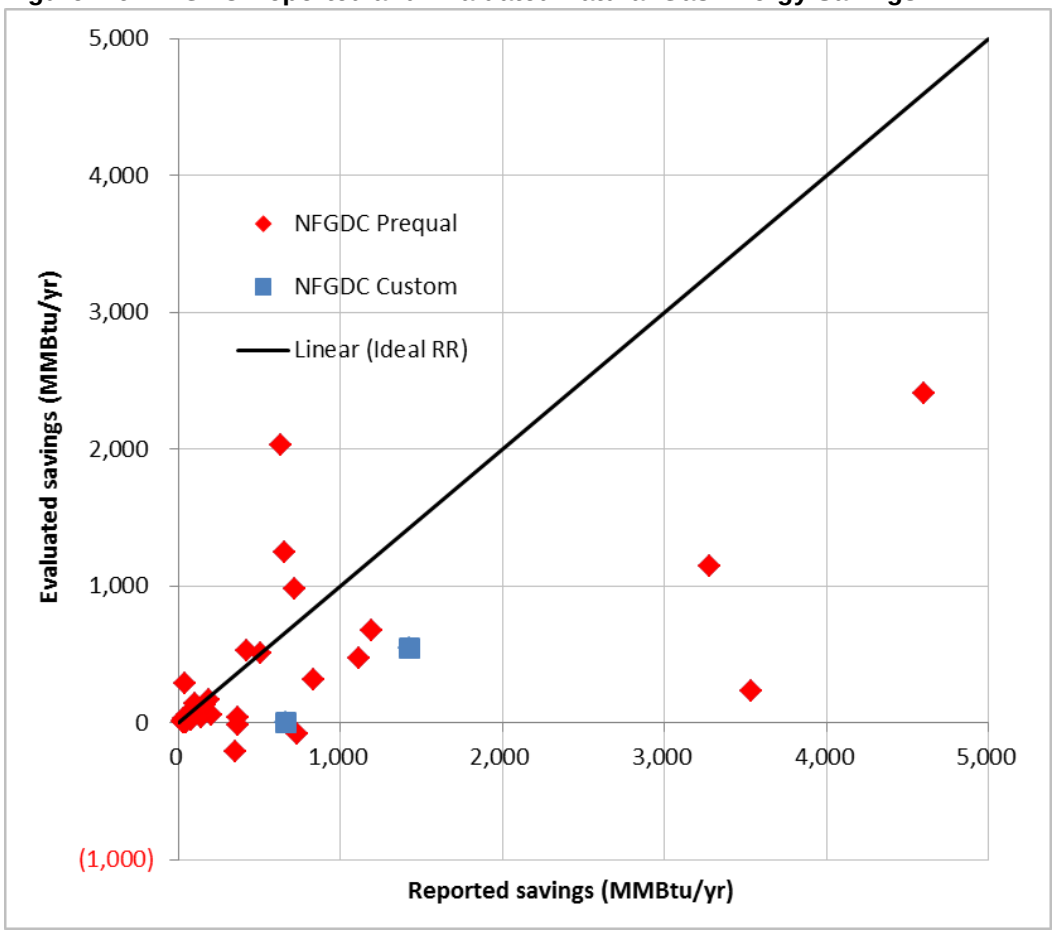
Table 4-8. NFGDC Program-Reported and Evaluated Natural Gas Energy Savings

Parameter	Program-Reported Savings	Realization Rate[1]	Evaluated Gross Savings	Relative Precision
Natural gas (MMBtu/yr)	33,137	0.66	21,870	26.2%

[1] Due to program changes that occurred in 2014, the Impact Evaluation Team recommends that NFGDC use an RR of 0.80 as a prospective pro forma value for 2014 and later reporting, until the next evaluation is completed.

Figures 4-6 illustrates the evaluated annual natural gas energy savings compared with those reported by the Program. Ideally, the evaluated gross savings would always match the program-reported savings. This ideal is shown as a solid black line on the charts. The actual findings are plotted as points on the graphs.

Figure 4-6. NFGDC Reported and Evaluated Natural Gas Energy Savings



4.3.2 Differences between Program-Reported and Evaluated Natural Gas Savings

For each project with an evaluated RR other than 1, a difference analysis was performed to identify the major driver or drivers to the evaluated RR. The results of this analysis are aggregated in an attempt to identify systematic differences to better inform Program staff. The differences were divided into primary categories and subcategories within the primary categories. The primary category indicated the phase of the project when the difference in the savings was likely to occur. The subcategories within the primary categories further provided detailed reasons for the difference in savings.

The results of the difference analysis are presented in Table 4-9. This table shows the positive and negative savings impact for the various differences. Differences that increased RR are shown as the blue bars, and differences that reduced RR are shown as the red bars.

Table 4-9. NFGDC Differences Analysis Results – Natural Gas

Discrepancy Category	Discrepancy Sub-Category	Counts	Positive and Negative Impact on RR*	Overall Impact on RR*	Total MMBtu Impact
Application review	Difference in as-built equipment efficiency	5			-304
	Difference in equipment hours of operation	5			3,344
	Fuel switching	1			-1,032
	Inaccurate estimation from applicant model	15			-1,906
	Inaccurate pre-project characterization	5			-1,267
	Incorrect baseline reference	3			-682
	Ineligible measure	1			-2,835
	Unknown applicant algorithm or assumptions	5			-3,021
Measure installation verification	Difference in installed control strategy	1			-420
	Difference in installed equipment size	1			8
	Difference in quantity installed	2			-407
	Inoperable measure	2			-2,177
Measure performance	Difference in equipment hours of operation	19			-8,800
	Difference in equipment load profile	3			7,896
	Difference in installed equipment efficiency	4			60
Total reported savings = 33,137 MMBtu					

* The blue bars indicate positive savings whereas the red bars indicate negative savings.

Table 4-9 above also illustrates the key contributors to the 0.66 RR for natural gas energy savings. Some of the major drivers to RR are discussed below:

- The application review category has the largest impact on RR. This category includes differences in baseline efficiencies, pre-project characterization, and installation of ineligible measures.
    - Systematic error in the prequalified savings calculation tool was identified to be the largest reason for lower RR for this category. The tool used an incorrect input to calculate the savings.
    - In addition, differences in pre-project characterization resulted in reduced RR for this category.
    - However, a difference in the equipment operating hours was found to improve the RR in this category.
- Measure installation verification category is the second-largest difference category driving down the RR. This category includes differences in installed equipment quantity, installed technologies, observed operations, and control strategies.
    - Within this category, two measures were found to be inoperable that influenced the RR to go down.

Appendix D[11] includes a detailed analysis of NFGDC natural gas RRs by measure type. Boiler and furnace efficiency improvements were the most common category and had relatively good RRs, an average of 87%. Space heating measures other than new boilers and furnaces had RRs averaging 34%.

4.4 OVERALL GAS AND ELECTRIC PERFORMANCE

NYSERDA EFP's electric program conspicuously outperformed both NYSERDA's and NFGDC's natural gas programs in terms of RRs, with RRs being about 40% higher for electric, closer to 1.0, and less variable. The natural gas program is new. The deemed savings have yet to be refined by years of experience in New York. Furthermore, natural gas measures are inherently harder to estimate, as it is expensive – and sometimes not even possible – to install short-term natural gas submeters. Savings projections tend to depend on proxy measurements. When measurements are available they often depend on the ability to install them during the right season, which doesn't always coincide with funding and application cycles. Pre-/post-retrofit billing analysis, a staple evaluation method for natural gas, is not an option for applicants. In short, there are many reasons for gas RRs to vary more than electric ones. Still, the magnitude of variation and overestimation for gas projects is significant and improvement should be possible as the programs mature and all parties gain more experience with natural gas measures.

4.5 SUMMARY OF 2010–2013 REALIZATION RATE

As described in the methodology, the Impact Evaluation Team developed a custom electric RR estimate for NYSERDA's 2010–2011 period by using weighted average site M&V-based RRs from the prior and subsequent evaluation periods and applied prior prequalified electric RRs directly. Table 4-10 summarizes the sources for the calculations.

Table 4-10. Source of Realization Rates by Installation Time Period and EFP Program

	Basis of Realization Rate Estimate For Projects Installed During the Period		
EFP Program	**1/1/2006 – 12/31/2009**	**1/1/2010 – 12/31/2011**	**1/1/2012–9/30/2013**
Electric prequalified	Prior evaluation	Prior evaluation	Prior evaluation
Electric custom	Prior evaluation	Blended RR	Current evaluation
Natural gas prequalified and custom	*Program did not exist.*	*Program did not exist.*	Current evaluation

The NYSERDA 2010–2011 period produced an electric RR of 1.02; there was no gas program.

[11] Results by building type were combined for both NYSERDA and NFGDC and are presented in Appendix C.

Table 4-11. Impact Evaluation Summary 2012–2013

Parameter	Program-Reported Savings (1/1/12 – 9/30/13)[1]	Realization Rate	Evaluated Gross Savings[2]	Relative Precision
NYSERDA				
Electric energy (MWh/yr)	221,473	1.01[1,2]	223,688	N/A[3]
Electric demand (MW)	35.9	0.98[1]	35.2	N/A
Natural gas energy (MMBtu/yr)	81,448	0.64[4]	52,127	7.9%
NFGDC				
Natural gas (MMBtu/yr)	33,137	0.66[5]	21,870	26.2%

[1] The electric RRs are based on the weighted average of the custom electric measure evaluated RRs from this study's primary research and the prequalified electric evaluated RRs from the prior EF impact evaluation.

[2] Realization rate and evaluated savings are rounded.

[3] The relative precision associated with the custom electric project sample is 6.4%.

[4] Due to program changes that occurred in 2014, the Impact Evaluation Team recommends that NYSERDA use an RR of 0.63 as a prospective pro forma value for 2014 and later reporting, until the next evaluation is completed.

[5] Due to program changes that occurred in 2014, the Impact Evaluation Team recommends that NFGDC use an RR of 0.80 as a prospective pro forma value for 2014 and later reporting, until the next evaluation is completed.

4.6 FINDINGS AND RECOMMENDATIONS

The Impact Evaluation Team offers one finding and two recommendations based on the impact evaluation research. These findings and recommendations are based on the sampled projects which included custom electric projects, and both custom and prequalified gas projects.

Finding – no action required:

The Impact Evaluation Team found that the deemed savings for the prequalified natural gas furnace and boiler measures used an incorrect basis for projects reported complete during the evaluation period. This inflated the savings estimates and reduced the RR. Program staff fixed the deemed savings tool in 2014. The Impact Evaluation Team reviewed the revised tool and found it to be accurate. No further action is recommended of Program staff. To the extent an RR is to be applied prospectively or on a pro forma basis, use 63% for NYSERDA and 80% for NFGDC, since this systematic issue has been addressed. These values considered the adjustments to the RR due to revised tool.

Recommendations:

1. The Program estimates custom electric energy savings well.

The results of the past two evaluations have resulted in RRs in close proximity to 1, relative precision values within the range of the sample design, and low error ratios. Program staff and their third-party Outreach and Technical Reviewers have demonstrated that they are capable of estimating project savings with high levels of engineering rigor.

Recommendation: This Program need not be subject to a traditional comprehensive retrospective gross impact evaluation of electric custom savings in the next cycle.

2. Prequalified gas applications tend to have the most variation in realization rates.

Prequalified gas projects were found to have the most variability between the reported and evaluated savings, resulting in lower relative precision values. By analyzing the RR differences attributable to the prequalified projects, Program staff can isolate differences that can be fixed during the application review process. Pre-project characterization was found to be one of the most common reasons for differences. Program staff could start by focusing on capturing the pre-project information accurately subsequently resulting in fewer differences in savings values.

Recommendation: Put procedures in place that will help take a closer look at some of the differences and associated inputs. This will subsequently improve the prequalified project savings estimates. Specifically, the two biggest causes of underestimation of prequalified reported compared to the evaluated savings were differences in equipment load profiles and differences in equipment operating hours. The program staff should be more attentive to the screening of prequalified projects as they tend to have the most variations.

APPENDIX A: GLOSSARY OF TERMS

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

error ratio – In energy efficiency evaluation, the error ratio is a measure of the degree of variance between the program-reported savings estimates and the evaluated estimates. For a sample, the error ratio is:

$$er = \frac{\sqrt{\sum_{i=1}^{n} w_i \frac{e_i^2}{x_i^{\gamma}} \sum_{i=1}^{n} w_i x_i^{\gamma}}}{\sum_{i=1}^{n} w_i y_i}$$

where,

n is the sample size

w_i is the population expansion weight associated with each sample point i

x_i is the program-program-reported savings for each sample point i

y_i is the evaluated gross savings for each sample point i, the constant gamma, $\gamma = 0.8$ (typically), and the error for each sample point $e_i = y_i - bx_i$, where b is the program realization rate

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

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

IPMVP Option A – This M&V option involves the partial measurement of isolated equipment affected by the evaluated measure. Relevant equipment variables are spot-measured when possible or stipulated when necessary.

IPMVP Option B – This M&V option involves full measurement of the isolated equipment affected by the evaluated measure. No stipulations are allowed. Both short-term and continuous data monitoring are included under Option B.

IPMVP Option C – This M&V option involves the use of utility meters to assess the performance of a total building. Option C addresses measure impacts in aggregate, not individually, if the affected equipment is connected to the same meter.

IPMVP Option D – This M&V option involves the use of computer modeling to determine facility or equipment energy use. Option D requires calibration with actual utility consumption data for either the pre-project or post-project period.

population expansion weight – The total number of units in a population divided by the number of units in the sample.

realization rate (RR) – The ratio of the evaluated gross savings to the program-reported savings. The RR represents the percentage of program-estimated savings that the Impact Evaluation Team estimates as being actually achieved based on the results of the evaluation M&V

analysis. For example, the RR calculation for electric energy for a sampled project is shown below:

$$RR = \frac{kWh_{evaluation}}{kWh_{program}}$$

where,

RR is the realization rate

$kWh_{evaluation}$ is the evaluation M&V kWh savings (by evaluation M&V contractor)

$kWh_{program}$ is the kWh savings claimed by program

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

Relative precision is calculated as shown below. It must be expressed for a specified confidence level. The relative precision (*rp*) of an estimate at 90% confidence is given below:

$$rp = 1.645 \ \frac{sd(\mu)}{\mu}$$

where,

μ is the mean of the variable of interest

sd(μ) is the standard deviation of μ

1.645 is the *z* critical value for the 90% confidence interval

For the 90% confidence interval, the error bound is set at 1.645 standard deviations from the mean. The magnitude of the *z* critical value varies depending on the level of confidence required.

stratified ratio estimator (SRE) – An efficient sampling design combining stratified sample design with a ratio estimator. It's most advantageous when the population has a large coefficient of variation, which occurs, for example, when a substantial portion of the projects have small savings, and a small number of projects have very large savings. The ratio estimator uses supporting information for each unit of the population when this information is highly correlated with the desired estimate to be derived from the evaluation, such as the program-reported savings and the evaluated gross savings.

summer coincident peak demand period – For this evaluation NYSERDA defined the summer coincident peak demand period as the energy reduction during the hottest non-holiday summer (June through August) weekday during the hour ending at 5 p.m.

weighted savings – Weighted savings are the program reported or evaluated savings multiplied by the population expansion weights.

APPENDIX B: ADDITIONAL ELECTRIC ENERGY SAVINGS BREAKDOWN

The Impact Evaluation Team also reviewed the evaluated projects to determine if patterns could be identified or feedback could be provided about project performance based on measure type, building type, and geographic location. The resulting observations are presented below. The results are suggestive regarding program areas of strength and opportunities for refinement regarding measure savings estimation. These additional results breakdown are provided to Program staff for information purposes only and are not for direct use elsewhere. Category sample sizes generally are too small to measure statistically significant differences in realization rates (RRs). RRs for individual measure and building types with samples of less than 10 should be considered anecdotal.

Electric Energy Savings for Custom Projects by Measure Type

Lighting and variable frequency drives (VFDs) were the most common measure type evaluated. While the program-level electric energy RR is 1.01, the unweighted RR for lighting and VFD measures were 1.13 and 1.04, respectively. These two measures types were also the most common measure types evaluated in the previous evaluation.

In addition, some cooling and associated controls measure types were also evaluated. The unweighted RRs for these measures were found to be low. The primary reason for the low RR was incorrect baseline reference (such as use of pre-retrofit chiller instead of code-compliant chiller) and difference in load profiles. In addition, the cooling equipment was found to operate fewer hours than those considered in estimating the program-reported savings.

Table B-1 provides a summary of measures within the sampled projects.

Table B-1. NYSERDA Number of Evaluated Measures and Realization Rate by Measure Group

Measure Type	Number of Measures	Realization Rate
Controls/VFDs	16	1.04
Cooling	10	0.31
Lighting and lighting controls	51	1.13

Lighting measures constituted roughly 65% of the evaluated gross savings of the Program. VFD and cooling related measures accounted for roughly 23% and 8% of the total evaluated gross savings, respectively.

Electric Energy Savings for Custom Projects by Building Type

Overall, the RRs for projects in commercial facilities (wholesale/retail, hospitality, office, services) tended to exceed those of non-profit facilities (education, not-for-profit, government). Commercial wholesale/retail and health care were the most common building types evaluated in both upstate and downstate regions. The unweighted RR for commercial wholesale/retail building type was found to be greater than 1 for both upstate and downstate regions. The relatively low RR for the education sector primarily is explained by the evaluated hours of use being lower than the applicant's. The predominant type of measures observed in the office buildings included lighting and VFD measures that had better RRs.

Table B-2 provides a summary of building types within the sampled projects.

Table B-2. Number of Evaluated Sites and Realization Rate by Building Type

Building Type	Number of Sites	Realization Rate[1]
Commercial – wholesale/retail	34	1.16
Education – colleges and universities	4	0.87
Education – elementary and secondary schools	3	0.96
Health care	8	0.79
Office	7	1.04

[1] Since RRs for individual measure and building types with samples of less than 10 have rather large confidence intervals, program staff should be cautious in making any significant changes in program targeting and delivery since the results are considered anecdotal.

APPENDIX C: ADDITIONAL NATURAL GAS SAVINGS BREAKDOWN FOR NYSERDA

The Impact Evaluation Team also reviewed the evaluated projects to determine if patterns could be identified or feedback could be provided about project performance based on measure type, building type, and geographic location. The resulting observations are presented below. While category sample sizes generally are too small to measure statistically significant differences in realization rates (RRs) and RRs for individual measure and building types with samples of less than 10 should be considered anecdotal, the results presented are useful for providing perspective on possible trends and feedback on project and measure performance.

Natural Gas Savings by Measure Type

Measures categorized as "gas efficiency" dominated the sample with more such measures (fifteen) than all other types combined. They include controls improvements such as thermostats and Energy Management System (EMS) and were found to have an unweighted RR of 0.82.

The furnaces, boilers, and space heating measure was the second-most common measure, and it was found to have unweighted RR of 0.61. This was primarily due to the systematic error[12] found in the deemed savings tool and difference in equipment load profiles for boilers and furnaces. This issue was fixed by Program staff in 2014. For space heating measures, the primary reason for change in savings was due to the difference in equipment load profiles (heating load requirements) followed by the difference in equipment efficiencies.

Table C-1 reflects the measure types evaluated on-site.

Table C-1. Number of Evaluated Measures by Measure Type – Natural Gas (NYSERDA)

Measure Type	Number of Measures	Realization Rate
Prequalified – furnaces/boilers/space heating	12	61%
Gas efficiency (controls, EMS, etc.)	15	82%

NYSERDA and NFGDC Natural Gas Savings by Building Type

The NYSERDA and NFGDC results were combined to take a look at the savings by building type with larger sample in each category. Educational institutions, commercial wholesale/retail, not-

[12] A systematic error in the furnace boiler prequalified savings calculation tool. The tool used the input boiler capacity where it should have used output capacity, resulting in overestimation of savings. However, this issue was fixed by Program staff in 2014 and all the projects after the tool fix used the correct algorithm.

for-profit, multifamily, and offices were the most common building types evaluated in the sample. The educational buildings were divided into universities and elementary/secondary schools. Offices, industrial, agriculture, health care, and hospitality categories all had realization rates of 57% or more whereas primary and secondary education, state government, and the not-for-profit categories all had realization rates of 37% or less, corroborating the pattern found for electric measures. Colleges and universities were an exception to this pattern.

Space heating measures in the schools were the primary reason for low RR. These measures had lower RR due to difference in efficiencies and operating hours. Industrial/manufacturing buildings were found to have RRs greater than 1. This was primarily due to higher furnace/boiler operating hours and installed efficiencies. The state government buildings primarily installed furnace/boiler measures. These measures had lower RRs primarily due to lower operating hours.

Table C-2 provides a summary of building types within the sampled projects.

Table C-2. Number of Evaluated Measures by Building Type – Natural Gas

Building Type	Number of Sites	Realization Rate[1]
Agriculture and forestry	1	82%
Commercial – wholesale/retail	8	32%
Education – colleges and universities	6	96%
Education – elementary and secondary schools	5	37%
Health care	2	87%
Hospitality	3	57%
Industrial/manufacturing	5	110%
Multifamily (over 4 units)	8	53%
Office	7	104%
State government	2	27%

[1] Since RRs for individual measure and building types with samples of less than 10 have rather large confidence intervals, program staff should be cautious in making any significant changes in program targeting and delivery since the results are considered anecdotal.

APPENDIX D: ADDITIONAL NATURAL GAS SAVINGS BREAKDOWN FOR NFGDC

The Impact Evaluation Team also reviewed the evaluated projects to determine if patterns could be identified or feedback could be provided about project performance based on measure type and building type. The resulting observations are presented below. Category sample sizes generally are too small to measure statistically significant differences in realization rates (RRs). RRs for individual measure and building types with samples of less than 10 should be considered anecdotal.

Natural Gas Savings by Measure Type

Furnace, boiler, and space heating efficiency measures dominated the sample with more such measures (31) than all other types combined. The low RR was primarily due to the systematic error[13] found in the deemed savings tool and difference in equipment load profiles. This error was corrected by Program staff in 2014.

Table D-1 reflects the measure types evaluated on-site.

Table D-1. NFGDC Number of Evaluated Measures by Measure Type – Natural Gas

Measure Type	Number of Measures	Realization Rate[1]
Prequalified – furnaces/boilers/ space heating	42	0.70
Prequalified – water heating	2	0.16
Gas efficiency	2	0.20

[1] Since RRs for individual measure and building types with samples of less than 10 have rather large confidence intervals, program staff should be cautious in making any significant changes in program targeting and delivery since the results are considered anecdotal.

Natural Gas Savings by Building Type

The NYSERDA and NFGDC results were combined to take a look at savings by building type. The Impact Evaluation Team combined the data provided the most value in analyzing the patterns (Appendix C).

[13] A systematic error in the furnace boiler prequalified savings calculation tool. The tool used the input boiler capacity where it should have used output capacity, resulting in overestimation of savings. However, this issue was fixed by Program staff in 2014 and all the projects after the tool fix used the correct algorithm.

APPENDIX E: EFP SAMPLE DESIGN MEMO FOR WINTER SITES

energy & resource solutions

120 Water St., Suite 350
North Andover, MA 01845
Phone: (978) 521-2550
Fax: (978) 521-4588
Web: www.ers-inc.com

MEMO

DATE: MARCH 05, 2014

TO: JUDEEN BYRNE, EVAN CRAHEN, AND ERIC MEINL

FROM: KATHRYN PARLIN, ISAAC WAINSTEIN, YOGESH PATIL, SUE HASELHORST, JON MAXWELL

RE: ON-SITE SAMPLE DESIGN AND M&V ACTIVITIES FOR EFP WINTER SITES (GAS ONLY)

The purpose of this memo is to present the on-site M&V sample design for NYSERDA Existing Facilities Program (EFP) projects with natural gas savings. The overall evaluation structure requires developing entirely separate samples for projects with natural gas savings and projects with electric savings. Since some of the natural gas projects require winter metering, this component of the EFP evaluation is on an accelerated schedule to allow the on-site survey to be fielded immediately. The New York Department of Public Service (DPS) has had an opportunity to discuss the overall EFP impact plan in a NYSERDA-DPS Strategy Session and a Sample Design Session (12/23/2013). This memo reflects the discussion of that session and subsequent discussions with NYSERDA and National Fuel Gas Distribution Corporation (NFGDC).

An impact work plan will be submitted shortly hereafter with additional plan details, including the sampling plan for the projects with electric savings. However, due to the urgency of fielding the on-site survey this winter, NYSERDA is requesting review and approval of the sample design for natural gas projects prior to submission of the entire work plan. The sample designs for attribution interviews and for the electric savings are not addressed in this memo, since the intent is to approve the on-site sample for projects with natural gas savings so that fielding can begin immediately.

EFP WINTER ON-SITE SAMPLE DESIGN FOR NATURAL GAS PROJECTS

For the verification of gross savings, the planned impact evaluation will include site survey work on a representative sample of participating facilities that have measures with natural gas savings. As noted in the DPS *Evaluation Plan Guidelines for EEPS Program Administrators*, the evaluation effort should be proportional to the uncertainty and level of program impact. These principles are used to consider tradeoffs in accuracy

and costs in sampling choices. Stratified ratio estimation will be used to improve precision and minimize sample sizes.

Establishing On-Site Sampling Precision Targets

The DPS guidelines require "90/10 confidence and relative precision for both net and gross saving at the program level. These requirements apply to each fuel, electric and gas." The DPS provided guidance in applying the propagation of error (POE) techniques[1] as a method of establishing the design precision for the components of the evaluation. DPS guidelines also recognize the need for cost effective, flexible evaluations, in that "a program might be so small that expending scarce evaluation dollars to achieve the 90/10 level of confidence and precision might not be cost-effective;" and that "some programs may be so small with commensurately small evaluation budgets that full compliance with these best practices may be impractical." This sentiment was further echoed by DPS Staff during the December Strategy Session.

Although this intention of this memo is to address only the on-site M&V sampling for projects with natural gas savings, the target precision was established by working backwards from the required precision for the overall net savings. This process required considering the POE by incorporating each component of the net natural gas savings.

The method for estimating the total sampling error associated with the net evaluated savings is a critical element in establishing sample sizes for the upcoming impact evaluations and will also influence data collection and analysis methods. The particular precision calculations required to compute the final program precision depends upon the following:

- The formulation of the equation used to compute net savings
- The magnitude of each of the terms in the equation
- The sampling precision of each term

The next section presents the results of the POE analysis of the net natural gas savings.

Propagation of Error Assumptions and Analysis

The current impact evaluation will include a sample of on-sites to determine the realization rate (RR); a sample of participants to be surveyed to determine free ridership (FR), inside spillover (ISO), and outside spillover (OSO); a sample of participating vendors to determine OSO; and finally the incorporation of the results from a separate study for the non-participant spillover (NPSO). The results from these efforts will be combined to compute net energy savings as follows:

Equation 1:

[1] *Evaluation Plan Guidance for EEPS Program Administrators* (*Evaluation Plan Guidance* or Appendix I)

$$Net\ energy\ savings_{MMBtu} = Reported\ Savings_{MMBtu} \times RR_{Rate} \times (1 - FR_{Rate} + NPSO_{Rate}) + OSO_{MMBtu} + ISO_{MMBtu}$$

This equation differs from NYSERDA's traditional net savings calculation formula, as shown below:

Equation 2:

$$\boldsymbol{Net\ energy\ savings_{MMBtu} = Reported\ Savings_{MMBtu} \times RR_{Rate} \times (1 - FR_{Rate} + ISO_{Rate} + OSO_{Rate} + NPSO_{Rate})}$$

The difference in the formulation of equation 1 is that ISO and OSO are directly estimated and added to the evaluated gross savings adjusted for FR and NPSO rather than being incorporated into the total net-to-gross ration (NTGR) as a percent of program savings.

Using equation 1 and assumptions from the previous EFP impact evaluation concerning the magnitude of the results and sampling precision, various precision scenarios were tested to arrive at the final on-site sample precisions presented in Table 1.

Table 1. Components of the Net Savings for Existing Facilities

Segment	MMBtu/yr	Estimated	Relative Precision	Notes
Tracking	173,293	N/A	N/A	Program-reported savings
RR	178,492	103%	**10%**	RR from prior EF impact evaluation for electric savings; natural gas was not evaluated. [Note that the FlexTech evaluation results showed a lower RR for natural gas (77% vs. 92% for electric measures)].
1- FR	119,572	69%	8%	Prior EF impact evaluation FR of 31% for electric measures. Precision is a target for (1-FR).
NPSO	43,323	25%	15%	Will use NPSO study finalized in December 2013.
ISO	20,795	12%	0%	Prior EF impact evaluation of 12% ISO for electric measures. Census planned.
OSO	55,454	32%	0%	Prior EF impact evaluation of 32% OSO for electric measures. Census planned.
Net energy savings	**244,031**	**141%**	**8%**	

The initial analysis suggested that an on-site M&V sample precision target of 12% would be adequate to result in a net energy savings precision of about 10%. However, this outcome assumes that the magnitude and precision of the components will be similar to the electric results from the previous impact evaluation conducted for PY2007/2008. There are several reasons where a more conservative approach is warranted, as explained below.

- There is uncertainty regarding the precision of all components listed in Table 1 above. For natural gas savings, the RR in other evaluations has been lower, and variability higher, in comparison to electric savings. Any decrease in the magnitude or precision of specific components will adversely affect the net savings precision.
- The net energy savings precision is particularly sensitive to the ISO and OSO estimates. If there is no SO, with all other components kept the same, the net savings precision erodes to over 12%. A reduction in SO magnitude (which may be likely, given the new SO procedures) negatively impacts the net savings precision.[2]

Based on the POE calculation and the uncertainty of the magnitude and precision of the factors (RR, FR, ISO, and OSO), the confidence/precision target for the program-wide on-site sample was set at 90/8 to allow some leeway in meeting the overall confidence/precision target for the net natural gas savings.

On-Site Sample Design

Each part of the sample design is described briefly in this section.

Sample Frame

The sample frame for on-site sampling includes all EFP participants who have completed projects from January 1, 2012, through September 30, 2013, and have natural gas savings. Similar to the previous evaluation, the sample frame was filtered to exclude the smallest projects accounting for the bottom 3% of savings for each segment.

Variable to Estimate

The purpose of this component of the evaluation is to estimate the RR and the gross evaluated savings. These savings will be multiplied by the net-to-gross (NTG) factors to obtain net savings.

Method

Stratified ratio estimation (SRE) will be used since it allows for efficient sampling design and generally requires a lower sample size for a targeted level of precision if there is a strong correlation between the program reported savings and the evaluated gross savings. SRE generally works well for estimating the RR because there usually is a strong correlation between these two variables.

Primary Sampling Unit

The sampling unit is the smallest combination of project and site. If a single site hosts two projects, each is a different sampling unit. Conversely, if a single project application

[2] A reduction in the magnitude of the SO reduces the total net savings. As the relative precision is calculated by dividing the error bound of the net savings by the total net savings, a reduction in the total net savings (the denominator) will result in worse precision if the error bound remains the same.

covers installation at two different sites, then each is a different sampling unit. The program tracking data are available for evaluators to systematically and uniformly make this separation. "Site" is defined as a single location that may include one or multiple buildings in close proximity (e.g. a premise). This approach was adopted to be able to implement the on-site M&V efficiently.

Error Ratio

The last Existing Facilities impact evaluation resulted in an electric measure error ratio of 0.50 and did not include gas measures. However, in order to get a sense of the error ratio for gas programs, we looked at the results from evaluations of custom gas programs in Massachusetts for Program Years 2009 and 2010. The evaluations for Program Years 2009 and 2010 showed a fairly dramatic improvement in the error ratio from 0.70 to 0.50, reflecting the maturation of the program.

The Impact Team concluded that the error ratio of 0.50 recommended for the electric program from the last evaluation is likely too optimistic for this fairly new program, and selected a design value of 0.60. This assumption will be reviewed based on the results of this evaluation. Because the precision target for estimating the sample size is more stringent than suggested by the POE analysis, the sample size should support a higher error ratio and still meet the confidence/precision target for the net natural gas savings. In addition, as the sampling for the NTG components is still pending, it may be possible to increase the sample sizes for FR and SO if the initial results of the on-site sample show higher variability than expected.

Upper-Level Stratification Variables

The upper stratification variables for the natural gas on site sample are described below.

1. Upstate vs. Downstate – All of the projects in Con Edison territory are considered as Downstate.
2. NYSERDA vs. NFGDC – The NFGDC program is administered by NYSERDA, but the savings are reported separately from NYSERDA's EFP.

Because NFG territory is exclusively upstate, this stratification plan leads to three strata: 1) Downstate NYSERDA projects, 2) Upstate NYSERDA projects and 3) Upstate NFGDC projects.

Lower-Level Stratification Variable

Size is used as the lower-level stratification variable. Size categories were based on the magnitude of the reported annual natural gas savings. Three size categories were defined. Cut-offs were established using the method described in the *2004 California*

Evaluation Framework.[3] The project size was defined as the total MMBtu gas savings at the site. The smallest projects accounting for 3% of the total savings were excluded for each stratum.

Sample Sizes by Stratum

The confidence/precision target was set for each stratum and the Impact Evaluation Team then verified that the targets for the strata result in a 90/8 overall target for the program-wide RR. The NFGDC component constitutes about 29% of the program gas MMBtu savings, with prequalified rebate measures representing 99% of NFGDC's full portfolio. Based on discussions with NYSERDA and NFGDC, the sample design for this component will target 90/10. Sample size calculations take in to account the finite population correction factor.

Table 2. Upstate NFGDC Sample Summary

Stratum	Number of Project /Sites	MMBtu Savings	% of Savings	Sample Size
0	78	985	3%	0
1	98	5,987	18%	8
2	24	10,185	31%	9
3	7	15,981	48%	7
Total	**207**	**33,138**	**100%**	**24**

The NYSERDA component comprises 71%of the total savings and is further stratified by Upstate and Downstate. The NYSERDA component excludes NFGDC projects. Both the Upstate and Downstate components have a target of 90/10. Sample size calculations take in to account the finite population correction factor.

Table 3. Upstate and Downstate NYSERDA Sample Summary

Stratum	Number of Projects	MMBtu Savings	% of Savings	Sample Size
Upstate				
0	28	1,537	3%	0
1	16	10,926	21%	5
2	7	14,721	28%	6
3	3	24,528	47%	3
Total for Upstate	**54**	**51,712**	**100%**	**14**
Downstate				
0	5	36	1%	0

[3] TecMarket Works, et. al. *The California Evaluation Framework*. Project Number: K2033910. Prepared for the California Public Utilities Commission and the Project Advisory Group. June, 2004. Pages 327 to 339 and 361 to 384.

1	5	6,024	20%	3
2	2	3,790	13%	2
3	2	19,560	66%	2
Total for Downstate	**14**	**29,736**	**100%**	**7**

Summary

The components of the on-site sample are summarized in Table 4.Table 4. Summary of On-Site Sample Components

Sample Component	Sample Approach	Comments
Sample frame	Projects completed from January 1, 2012 through September 30, 2013	Billing analysis is the preferred approach for majority of the prequalified measure types (wherever it is feasible).
Variable to be estimated	Evaluated gross MMBtu	Program-reported natural gas savings.
Method	Stratified ratio estimation	SRE generally works well for gross RR as there is usually a strong correlation between program-reported and evaluated savings.
Primary sampling unit	Project/Site (location)	A site is a location that may include a single building or multiple buildings in close proximity.
High-level stratification	1) Upstate NFGDC 2) Upstate NYSERDA 3) Downstate NYSERDA	Sample sizes were calculated to meet the precision/confidence target for each domain of interest.
Lower-level stratification variables	Size	Three size categories were defined.
Definition of size	Gas MMBtu savings	Cut-offs were established using the method described in the CA Frameworks.
Expected precision of the on-site sample	Upstate and Downstate	10% with confidence at the 90% interval
	NFGDC	10% with confidence at the 90% interval

Table 5 shows the sample sizes and target precision for each component of the EFP. These sample sizes are based on using SRE and take in to account the finite population correction factor and the precision advantage of including a census of large projects in the sample.

Table 5. Summary of On-Site Sample Sizes by Program Component

Program Component	Total Number of Projects	% of Reported Savings, MMBtu	Random Sample	Census Projects	Total Sample Size	Target Precision of On-site Sample	Expected Range of Net Savings Precision
NFGDC	207	29%	24	0	24	90/10	8-12%
Upstate	54	45%	11	3	14	90/10	8-12%
Downstate	14	26%	5	2	7	90/10	8-12%
Statewide totals	**275**	**100%**	**40**	**5**	**45**	**90/8**	**6-9%**

This strategy provides a scenario where it is likely the net savings precision target of 10% will be met on the upstate/downstate domains and a worst-case scenario where the evaluation guideline precision target is met statewide. NFGDC's program represents 29% of the total savings being evaluated. NFGDC, NYSERDA, and ERS have participated in project scoping meetings between December 2013 and February 2014, to initiate the development of a more detailed evaluation work plan and approach for a collaborative joint evaluation. Part of this discussion has focused on the logistics of delivering a statistically significant evaluation product in a cost-effective manner.