Itron
Knowledge to Shape Your Future

Electric | Gas | Water
information collection, analysis and application

2010-12 WO033 Custom Impact Evaluation Final Report

Submitted to:

California Public Utilities Commission
505 Van Ness Avenue
San Francisco, CA 94102

Submitted by:

Itron, Inc.
1111 Broadway, Suite 1800
Oakland, CA 94607
(510) 844-2800

and

DNV GL
155 Grand Avenue, Suite 500
Oakland, CA 94612

With Assistance from:

Energy and Resource Solutions
Energy Metrics
Michaels Engineering
PWP, Inc.
Katin Engineering Consulting
Robert Thomas Brown Company
Leidos, LLC
Warren Energy Engineering, LLC

July 14, 2014

Abstract

This report presents the evaluation results for energy efficiency projects and programs within the scope of the custom impact evaluation conducted for the California Public Utilities Commission (CPUC) Energy Division (ED) under Work Order 33 (WO033). The evaluation addresses program impacts for the 2010–2012 program cycle and concentrates on non-deemed projects and programs in the commercial, industrial and agricultural sectors. Gross impacts, net impacts, and qualitative program assessment results are reported for the four main California IOUs. Five sampling domains were selected for the evaluation: PG&E Electric, PG&E Gas, SCE Electric, SDG&E Electric and SCG/SDG&E Gas.

The evaluation results are based on a gross impact sample of 429 projects and yield gross realization rates ranging from 0.59 to 1.40 across IOU fuel domains for the primary fuel evaluated. Gross realization rates range from 0.58 to 0.67 when the most extreme realization rates are excluded. Net to gross ratio (NTGR) results are based on 1,388 sample points and the resulting IOU fuel domain NTGRs ranged from 0.46 to 0.56. The overall net evaluation realization rates (evaluated savings compared to unadjusted gross claims) for the primary fuel evaluated vary from 0.28 to 0.70 across the five IOU fuel domains.

A novel element incorporated in this evaluation was a qualitative lower rigor assessment (LRA) for 536 energy efficiency projects. These LRAs used an engineering desk review of project files to assess programs and program groups. The strength and weaknesses of programs and program groups were assessed through 17 key metrics related to gross impact estimation practices. Key findings address the relative performance of core and non-core program offerings, including third party, new construction, and government partnership programs.

The evaluation examined the frequency and effects of various discrepancy factors on gross impacts. The greatest difference between ex ante and evaluation impacts were from changed operating conditions, errors in baseline determination, and changes to calculation methods; eligibility, measure count variance, equipment specification changes, and tracking system discrepancies affected the gross impact to a lesser extent. The evaluation also quantified the occurrence of factors responsible for low net to gross ratios (NTGRs). Common themes related to free ridership were that project adoption was often related to corporate policy or regulatory compliance, non-energy benefits, or decisions to implement prior to program application.

Stakeholders can use the findings of this evaluation to concentrate on key issues and factors that, if successfully addressed, can improve project gross and net savings estimation and better align claimed and evaluated savings.

Table of Contents

List of Figures

List of Tables

1

Executive Summary

This report presents findings from the custom project impact evaluation of the program year (PY) 2010-2012 California investor-owned utilities' (IOUs') energy efficiency programs.[1] This custom project impact evaluation is one of multiple California Public Utilities Commission (CPUC) evaluations of the IOUs' 2010-2012 efficiency programs. This custom evaluation was conducted under the Work Order 33 (WO33) as part a prime contract for PY2010-2012 evaluation services.

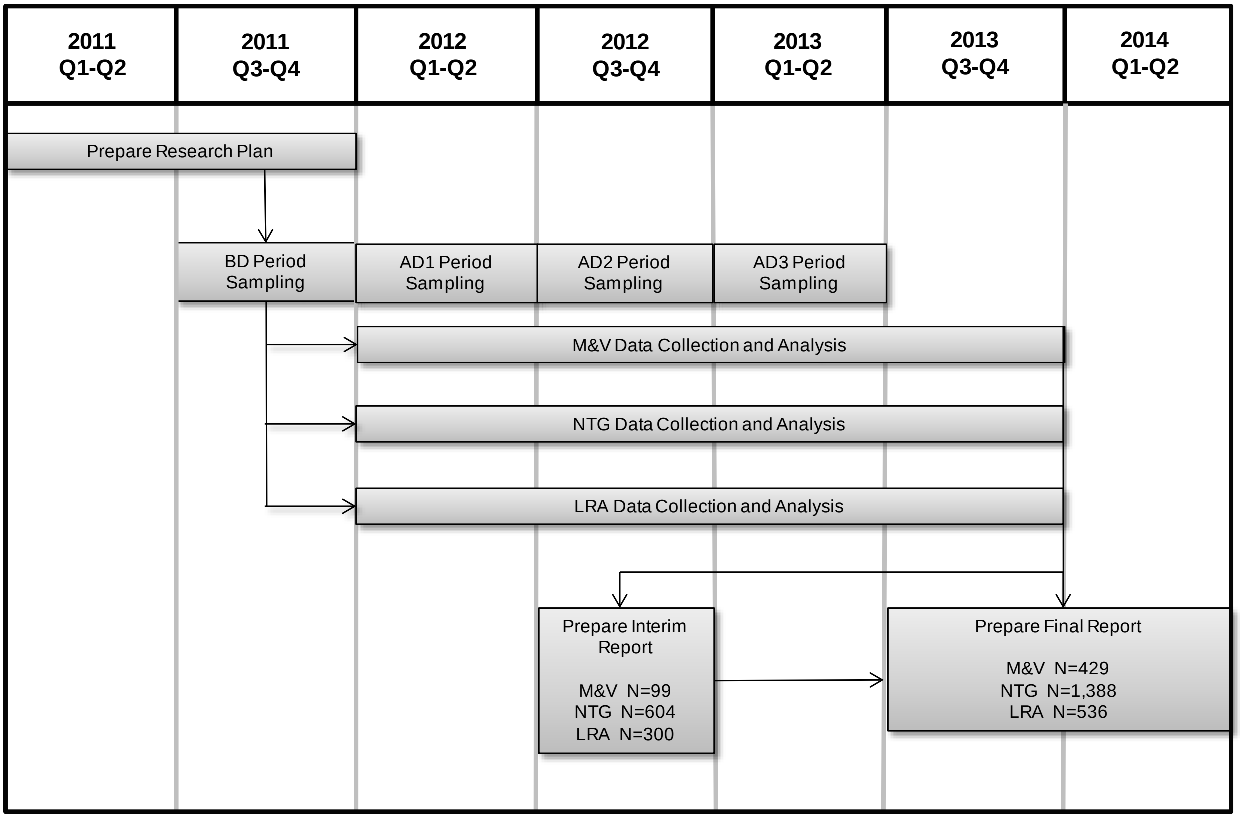
The evaluation addresses custom, non-deemed, measure installations, and involves an array of projects that received incentives via more than 100 utility programs. The IOU programs evaluated span all offerings where custom incentives are provided for non-deemed measure installations. The scope of work for the evaluation of custom measures includes an independent estimation of gross impacts and net impacts, and a lower rigor assessment (LRA) activity[2] to broaden the reach of the evaluation into programs that would not otherwise be examined due to sample size constraints associated with more resource intensive measurement and verification (M&V) used for estimated gross impacts. Extensive findings and recommendations to improve program and project effectiveness are also provided.

Three main evaluation activities support the findings and recommendations in this report: (1) M&V activities for estimating gross impacts for 429 projects across five IOU fuel analysis domains, (2) telephone survey data collection supporting net to gross (NTG) estimation for a total of 1,388 projects (by IOU fuel as well as key programs and program groupings), and (3) a total of 536 engineering reviews supporting the qualitative LRA and associated segment-level results. A high-level overview of the project timeline and key activities is shown in Figure 1-1.

1 This effort was completed for CPUC under the direction of staff responsible for evaluation of utility energy efficiency programs.

2 Gross impact projects and 'lower rigor only' projects are analyzed jointly to examine conformance with program procedures (including measure eligibility and other rules); analyze strengths and weaknesses of project applications; and provide feedback on program ex-ante impact estimates.

Figure 1-1: High Level Overview of Project Timeline and Key Activities



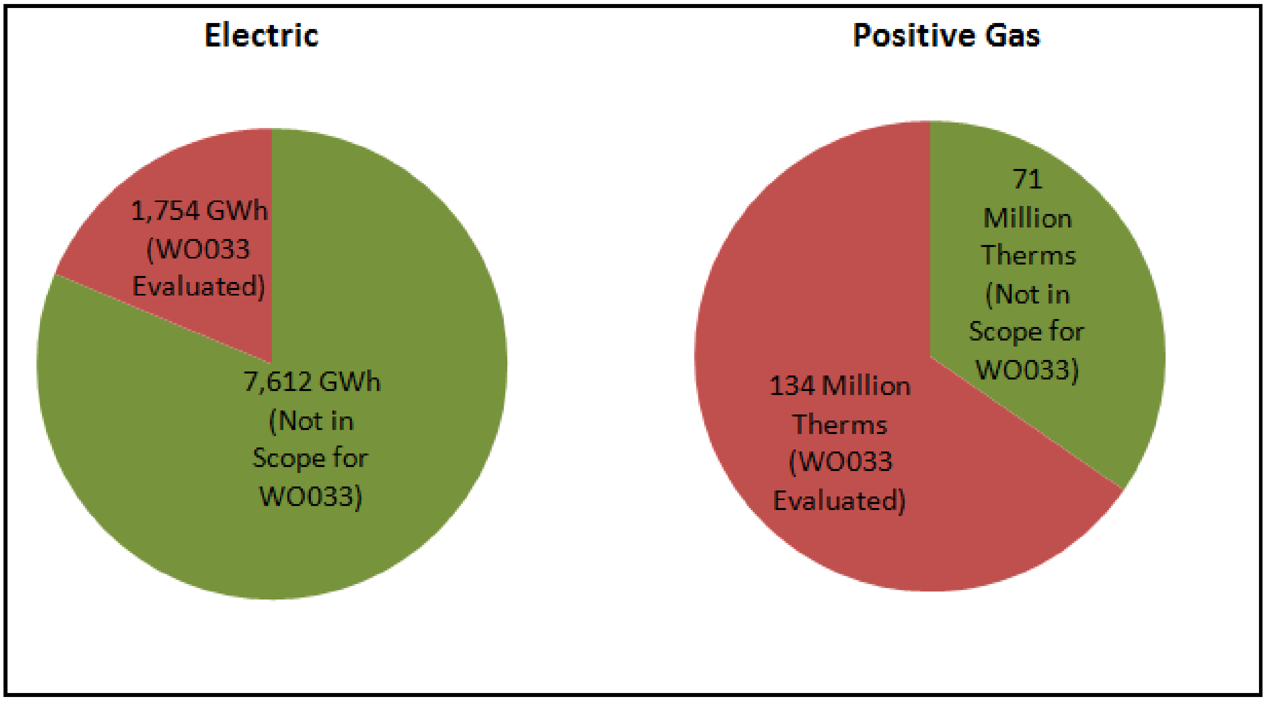
1.1 Custom Impact Evaluation Portfolio Context and Sample Sizes

The programs included in this custom impact evaluation primarily address industrial and manufacturing facilities; water supply, water treatment and wastewater treatment; oil and gas extraction, oil refining and production; and commercial custom, non-deemed program offerings (including the Savings by Design new construction program and retro-commissioning programs). The scope addresses nonresidential custom measures of all types with one exception: lighting measures were generally excluded, except where a project's scope addressed the whole building (e.g., commercial new construction projects).[3] Each custom-oriented program offers one or more of the following interventions in order to encourage end users to upgrade to energy-efficient measures: site specific facility assessments, feasibility studies, project incentives, facility audits, pump testing, and specialized training.

[3] Custom lighting measures are addressed in a separate impact study on nonresidential lighting, the CPUC 2010-2012 evaluation Work Order 29.

As shown in Figure 1-2, energy savings claims associated with the scope of this evaluation represent a significant contribution to the overall savings portfolios for the IOUs' energy efficiency programs, accounting for about 19 percent of statewide electric savings claims and 65 percent of statewide gas savings claims[4] during 2010-12. During this period, the IOU tracking data for measures associated with this custom impact evaluation included thousands of entries statewide with annual electric savings claims by the IOUs totaling 1,754 GWh. Statewide IOU annual gas savings claims for measures included in this impact evaluation totaled 134 million therms.

Figure 1-2: Custom Impact Evaluation Share of Statewide PY2010-2012 Energy Efficiency*



* "Positive" gas refers to exclusion, in this view, of negative gas claims associated with the interactive effects of electric measures (e.g., lighting).

A variety of possible sampling domains were considered for this evaluation, including dimensions such as programs and measure groups. IOU and fuel were ultimately chosen as the primary sampling domains for developing and reporting gross impact results due to the number of gross impact M&V sample points targeted for the study and the number of sample points required to provide reasonable statistical precision for a sampling domain. This approach resulted in the following five sampling domains for which gross realization rates were developed and reported: PG&E electric, PG&E gas, SCE electric, SCG/SDG&E gas, and SDG&E electric. Because of the lower costs per sample point associated with the NTG and LRA evaluation

4 Excluding negative claims associated with HVAC interactive effects.

activities, larger sample sizes were used for those research tasks. This resulted, by design, in the ability to report results at more detail than IOU fuel, for example, NTG and LRA results are also reported by program group. The total sample sizes for each activity are shown in below.

Table 1-1: Summary of Custom Evaluation Sample Sizes for IOU – Fuel Domains

Utility/Fuel Sampling Domain	Number of Completed Surveys (n)		
	Gross Impact (M&V)*	**NTG**	**LRA**
PG&E Electric	112	558	155
PG&E Gas	75	230	97
SCE Electric	100	367	139
SDG&E Electric	73	125	73
SDG&E/SCG Gas	69	108	72
Total	**429**	**1,388**	**536**

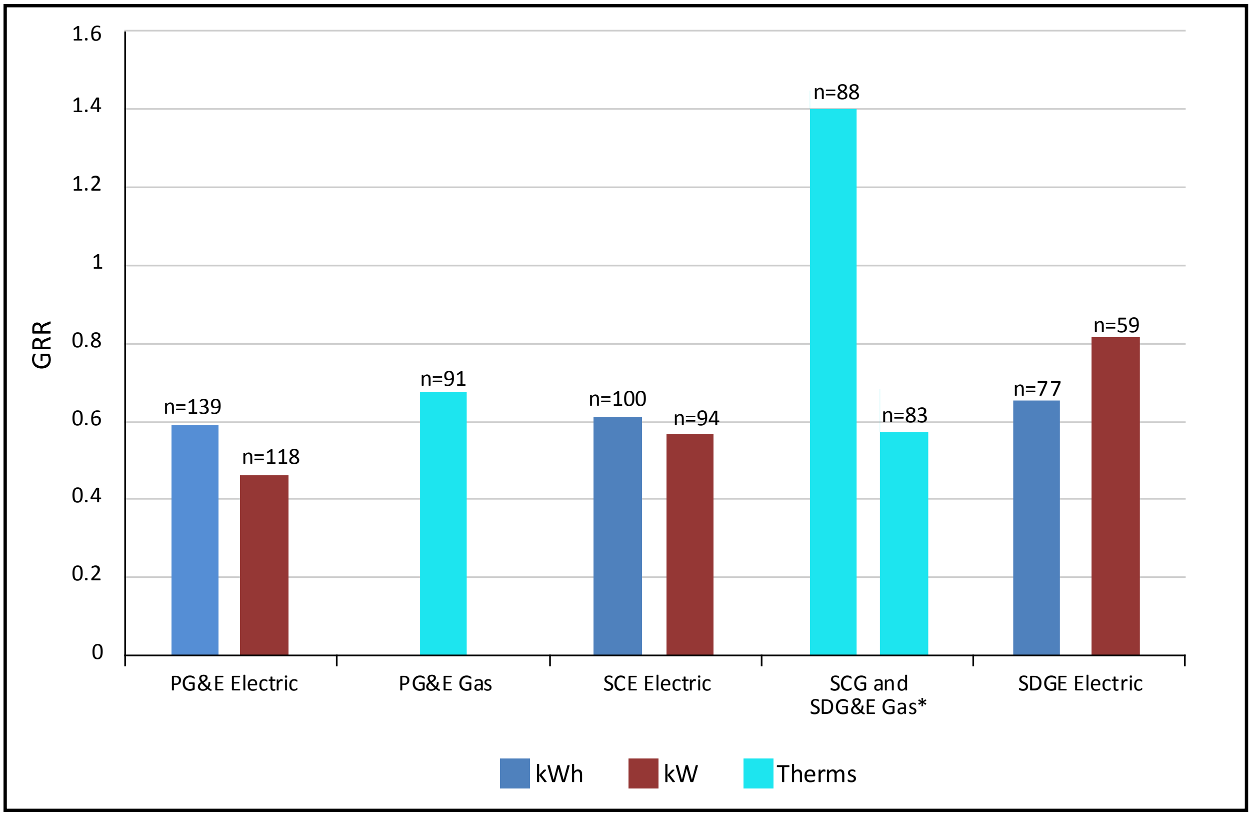
* The 429 M&V sample points account for 495 electric and gas fuel-differentiated results. This is because some sample points included claims for both electric and gas savings. These 495 results are referred to as "projects."

1.2 High-Level Custom Gross Impact Results

In Figure 1-3 and Table 1-2 below, we summarize the mean gross impact realization rates (GRRs) for each of the five IOU fuel sample domains. Realization rates are calculated for each sampled project as the ex-post, evaluation based estimate of impacts divided by the IOUs' ex-ante estimate of impacts. Sample weights are used to extrapolate the evaluation results to the population. The population sample frame and the total number of completed gross impact points are also shown in Table 1-2 for each energy metric, along with the resulting error ratio (ER), which is a measure of the statistical variation in the gross realization rates, and the 90 percent confidence intervals. With all sample points included, the mean realization rates by IOU, fuel and energy metric are less than 0.70 for all but two energy metrics. One domain, SCG/SDG&E gas, has a GRR that exceeds one, while SDG&E kW has a GRR of 0.82. The results for the SCG/SDG&E gas domain are strongly impacted by results from three projects with extreme GRRs of 40, 46, and 53. All were sampled as SDG&E *electric* projects, but had relatively small ex-ante, secondary gas impact claims that were estimated to be significantly larger in the ex-post evaluation. Gross realization rates with extreme points removed result in a GRR of 0.58 for the SCG/SDG&E gas domain.

The mean gross realization rates are in some cases higher, lower, or relatively equal to results from the CPUC's evaluation of the IOUs' 2006-2008 custom programs; generally, they remain significantly below the 0.9 default ex-ante GRR adjustment for custom programs.[5] A significant number of projects were estimated to have negative and/or zero GRRs in all domains.

Figure 1-3: Project Lifecycle Gross Realization Rates by Sample Domain and Energy Metric (kWh, kW, and Therms)



* Lower value excludes extreme points.

5 With the exception of SCG/SDG&E gas depending on treatment of the secondary fuel outliers.

Table 1-2: Project Lifecycle Gross Realization Rates by Sample Domain and Energy Metric (kWh, kW, and Therms)

Energy Metric	Sample Size (n)	Mean Gross Realization Rate	Population (N)	Error Ratio***	90% Confidence Interval
PG&E Electric					
kWh*	139	0.59	6,994	1.26	0.49 - 0.70
kW	118	0.46	6,248	1.67	0.35 - 0.58
PG&E Gas					
Therms*	91	0.67	1,270	0.56	0.61 - 0.74
SCE Electric					
kWh*	100	0.61	3,052	1.03	0.51 - 0.71
kW	94	0.57	2,748	1.08	0.47 - 0.67
SDGE Electric					
kWh*	77	0.64	1,469	0.99	0.52 - 0.75
kW	59	0.82	790	2.10	0.46 - 1.17
SCG and SDG&E Gas					
Therms, all points*	88	1.40	1,077	4.99	0.23 - 2.57
Therms, less extreme points**	83	0.58	1,077	0.80	0.50 - 0.66

* The primary sample was designed and selected at this level. The kW sample sizes are lower due to the fact that kW impacts were not claimed by IOUs in every case.

** The SCG/SDG&E gas domain was strongly affected by two extreme points associated with underestimation of secondary fuel (gas) impacts in two SDG&E electric sample points. Results are shown excluding these and two other extreme cases (i.e., excluding a total of four points).

*** A measure of the statistical variation in the gross realization rates.

The three principal reasons that ex-ante gross impacts differ from ex-post results are: (1) observed operating conditions, (2) baseline specification, and (3) the IOU's calculation methods. These discrepancy factors explain a portion of the differences in ex ante savings and ex post results for 45 percent, 17 percent and 20 percent of the records studied.[6] Across all IOUs, the respective impact of these three factors on electric savings are roughly -17 percent,-16 percent and -2 percent, respectively; while their effect on gas savings claims are roughly -13 percent, -14 percent and +1 percent.[7]

In an effort to provide more specific and actionable findings, GRR results for program and measure groups with substantial sample sizes were examined in detail. Differences were found between core, third party and new construction programs; core groups and some third party

[6] Other reasons for differences in savings results were observed less frequently, but include the following: incorrect equipment specifications, ineligible measures, incorrect measure counts and tracking database discrepancy, among others.

[7] Gas discrepancy effects exclude the two SDG&E extreme GRRs for secondary fuel impacts mentioned above.

programs were found to have relatively high gross impact results. The performance of measure groups was also examined. While quite sensitive to outliers, lower GRR results were concentrated in the compressed air, HVAC and HVAC control measure groups. Complete gross impact results developed for program and measure groups are presented in Chapter 5 and Appendix C.

1.3 High Level Custom Net-to-Gross Results[8]

NTG results at the level of sampling domain (IOU fuel) are presented in Figure 1-4 and Table 1-3. Evaluation net-to-gross ratio (NTGR) results reveal that significant free ridership has persisted into this program cycle. On a Statewide basis, the NTGR across all program categories averaged 0.48 for electric programs and 0.53 for gas programs. These values indicate a medium high[9] level of free ridership, and a resulting medium low level of program influence. Evaluated NTGRs are similar in magnitude to those from the results of evaluations dating back to program year 1998.

While there are many root causes of high free ridership present in this market segment (the large industrial market, in particular, can be very difficult to influence in practice),[10] we found little evidence of any specific actions taken to change program design features, requirements, procedures, guidelines, resources, or incentives to try to reduce free ridership for custom projects. Recommendations for reducing free ridership in custom programs have been provided many times in previous evaluation reports, including the CPUC's PY2006-2008 evaluation studies, as well as IOU managed evaluation studies dating back to the late 1990s and early 2000s. We observed limited evidence of significant free ridership-related changes in custom program design and implementation procedures for the 10-12 program cycle; however, we note that PG&E reported in its comments on the draft of this report that they have made important changes.[11] The CPUC and evaluators do not believe these changes occurred early enough to affect the projects evaluated for PY2010-2012. Evidence for these changes, both qualitative and quantitative should be a focus of subsequent program year evaluations.

---

8 In addition to the NTG results in this custom impact evaluation report, a separate report focusing on more detailed custom NTG analysis methods, data sources, findings and recommendations. This report will be made available for review in late March 2014.

9 Defined as a free ridership level of between 50 percent and 74 percent (i.e., an NTGR of between 0.26 and 0.50).

10 This is due principally to the highly technical, industry- and site-specific process equipment requirements of these firms and the high fraction of energy use as a cost of production for some segments. This results in a strong internal incentive, as well as availability of internal expertise and resources, to maximize production efficiency.

11 The changes cited by PG&E include: 1) PG&E has increased outreach efforts to hard-to-reach small and medium businesses; 2) PG&E has improved its program influence project documentation and trained Customer Relationship Managers on the importance of capturing such data; 3) PG&E is carrying out Industry Standard Practice (ISP) studies to identify measures for sun-setting from their portfolio including pump-off controllers in 2009.

Figure 1-4: Weighted Net-to-Gross Ratios by IOU Fuel Domain[12]

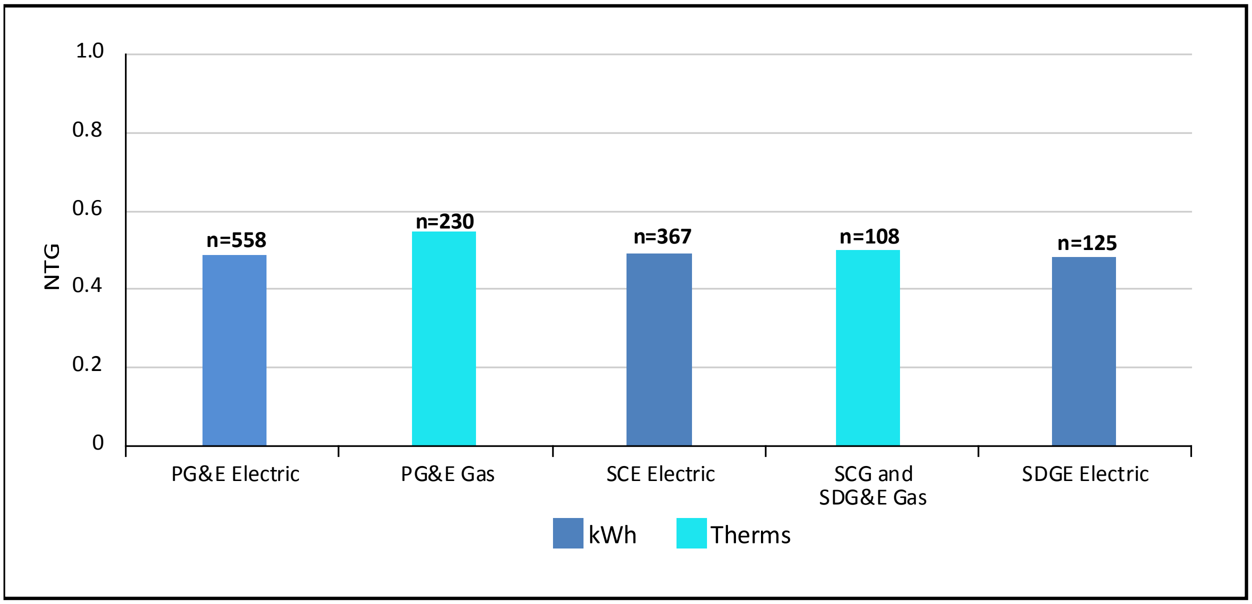


Table 1-3: Weighted Net-to-Gross Ratios by IOU Fuel Domain[13]

Results	Electric NTGRs				Gas NTGRs		
	Statewide	PGE	SCE	SDG&E	Statewide	PGE	SDG&E/SCG
Weighted NTGR	**0.47**	**0.46**	**0.49**	**0.45**	**0.53**	**0.56**	**0.50**
90 Percent Confidence Interval	0.46 - 0.48	0.45 - 0.48	0.47 - 0.50	0.43 - 0.48	0.50 - 0.57	0.53 - 0.58	0.42 - 0.57
Relative Precision	0.03	0.04	0.04	0.06	0.07	0.05	0.15
n NTGR Completes	1,050	558	367	125	338	230	108
N Sampling Units	11,515	6,994	3,052	1,469	2,347	1,270	1,077
Error ratio (ER)	0.52	0.59	0.44	0.43	0.79	0.46	0.99
NTGR Adjustment Factor	1.02	1.01	1.02	1.02	1.00	1.00	1.00
Final NTGR	**0.48**	**0.47**	**0.50**	**0.46**	**0.53**	**0.56**	**0.50**

12 Note that these values do not include the effects of 9 projects for the limited purpose of calculating an NTGR Adjustment Factor. As described in Chapter 4, this was due to the potential overlap for these sites between the NTG for the project and the Gross ISP or Dual Baseline determinations in the gross impact analysis. For all IOU-fuel domains except PG&E and SDG&E/SCG Gas, the improvement in IOU-fuel domain level NTGRs from these removals was very slight, on the order of 1 to 2%. The PG&E Gas value had no projects removed and therefore, remains unchanged. The change in the SDG&E/SCG Gas NTGR value was extremely low, resulting in a 1.00 multiplier.

13 See previous footnote re NTG adjustment factor.

More detailed results, at the program and program group levels are presented in Chapter 6.[14] The key findings from this more detailed analysis are:

- ***Certain market segments and energy efficient measure categories have higher estimated free ridership***. For example, within the Water-Wastewater sector, NTGRs for the Water Supply/Irrigation and Sewage Treatment Facilities categories were among the lowest, with values of 0.29 and 0.27, respectively. Among the measures analyzed, both Aerators and Controls had particularly low NTGRs. VFD NTGR results, especially for PG&E, also demonstrated a low level of program influence. Similarly, for the Agricultural sector, the Agriculture and Forestry category had among the lowest NTGR levels (0.39 for PG&E Agriculture and Forestry; 0.41 SCE Water Supply Irrigation).
- ***Estimated net-to-gross ratios for certain categories analyzed are lower as compared to the PY2006-2008 evaluation***. For Retrocommissioning programs, the current NTGR findings for SCE of 0.61 (15 projects) and for PG&E of 0.58 (47 projects) are substantially less than the PY2006-2008 evaluation values of 0.75 and higher for all IOU programs results.

Behind the NTGR calculated for each project are a host of contextual factors that may have influenced the project, directly or indirectly. The key contextual factors were first examined within each project, and then summarized across all evaluated projects within a given program or program grouping. The intent was to look more deeply, beyond the numerical responses used in the NTGR algorithm, into the qualitative factors that influenced the project decision making. Across all programs and program groupings analyzed, corporate policy was a major driver for most projects. Related to this was the presence of corporate policy associated with environmental protection. For programs and program groups with the lowest NTGRs, there were one or more other strong drivers present that contributed to reduced program influence. For SCE, a common theme was replacement of failing equipment. For SDG&E, environmental compliance features prominently. For some PG&E projects, additional non-energy benefits like automation were cited as the project driver, and low program influence was evident when projects were already in advanced stages of design and implementation (and therefore not influenced substantially by the program). Finally, for new construction projects, a significant percentage of projects were implemented by firms already using advanced energy efficiency in designs, including national chains and big box stores.

Corporate policies that favor energy efficiency investment are a favorable characteristic of end-user decision making that aligns well with the state's overall, long-term goals for energy efficiency adoption and climate change mitigation. At the same time, correlation of this and other efficiency oriented decision making characteristics with program free ridership presents

[14] For categories where there were sufficient numbers of completed surveys; generally eight or more.

challenges to program designers, administrators and implementers with respect to policy objectives to maximize the net impact of ratepayer-funded efficiency programs while also achieving aggressive total savings goals.

1.4 Net Evaluation Realization Rate Results

Net evaluation realization rates are presented in Table 1-4 through Table 1-6 by IOU fuel domain. Net realization rates are the product of the GRRs and the NTGRs, and thus portray the combined evaluation impact as compared to unadjusted IOU ex ante gross impact claims. These tables also provide a comparison of the ex post net realization divided by the IOUs' net realization rates (that is, the evaluation results compared to the IOUs' ex ante values inclusive of the default GRR of 0.9 and the IOUs' ex ante NTG values).

Table 1-4: PG&E Net Realization Rate Estimates and Comparisons

	Electric Savings		Gas Savings
Impact Element	**kWh/year**	**Average Peak kW**	**Therms/year**
Tracking			
a. Claimed Gross Savings	1,013,623,875	143,553	84,070,206
b. Claimed GRR	0.90	0.90	0.90
c. Claimed Adjusted Gross Savings (c = a x b)	912,392,658	129,222	75,882,422
d. Claimed NTGR	0.70	0.70	0.67
e. Claimed Net Savings (e = c x d)	639,455,515	90,181	50,883,796
f. Claimed Net Realization Rate (f = b x d)	0.63	0.63	0.61
Evaluation			
g. Evaluation LC GRR	0.59	0.46	0.67
h. Evaluated Gross Results (h = a x g)	600,295,707	66,465	56,731,743
i. Evaluation NTG Ratio	0.46	0.46	0.56
j. Evaluation NTG Ratio Adjustment	1.01	1.01	1.00
k. Evaluated Net Results (k = h x i x j)	281,834,111	31,205	31,582,214
l. Evaluation Net Realization Rate (l = g x i x j)	0.28	0.22	0.38
m. Evaluated Net Savings as a Fraction of Claimed Net Savings (m = l / f)	0.44	0.35	0.62

Table 1-5: SCE Net Realization Rate Estimates and Comparisons

Impact Element	Electric Savings	
	kWh/year	Average Peak kW
Tracking		
a. Claimed Gross Savings	738,229,082	107,988
b. Claimed GRR	0.90	0.90
c. Claimed Adjusted Gross Savings (c = a x b)	664,476,265	97,195
d. Claimed NTGR	0.67	0.67
e. Claimed Net Savings (e = c x d)	446,687,086	65,365
f. Claimed Net Realization Rate (f = b x d)	0.61	0.61
Evaluation		
g. Evaluation LC GRR	0.61	0.57
h. Evaluated Gross Results (h = a x g)	448,068,199	61,365
i. Evaluation NTG Ratio	0.49	0.49
j. Evaluation NTG Ratio Adjustment	1.02	1.02
k. Evaluated Net Results (k = h x i x j)	222,373,298	30,455
l. Evaluation Net Realization Rate (l = g x i x j)	0.30	0.28
m. Evaluated Net Savings as a Fraction of Claimed Net Savings (m = l / f)	0.50	0.47

Table 1-6: SDG&E/SCG Net Realization Rate Estimates and Comparisons

Impact Element	Electric Savings		Gas Savings
	kWh/year	Average Peak kW	Therms/year
Tracking			
a. Claimed Gross Savings	196,589,992	26,725	62,461,839
b. Claimed GRR	0.90	0.90	0.90
c. Claimed Adjusted Gross Savings (c = a x b)	176,930,993	24,052	56,239,890
d. Claimed NTGR	0.69	0.74	0.65
e. Claimed Net Savings (e = c x d)	122,468,747	17,782	36,304,928
f. Claimed Net Realization Rate (f = b x d)	0.62	0.67	0.58
Evaluation			
g. Evaluation LC GRR	0.64	0.82	1.40
h. Evaluated Gross Results (h = a x g)	124,991,016	21,871	87,213,930
i. Evaluation NTG Ratio	0.45	0.45	0.50
j. Evaluation NTG Ratio Adjustment	1.02	1.02	1.00
k. Evaluated Net Results (k = h x i x j)	58,103,201	10,167	43,572,259
l. Evaluation Net Realization Rate (l = g x i x j)	0.30	0.38	0.70
m. Evaluated Net Savings as a Fraction of Claimed Net Savings (m = l / f)	0.47	0.57	1.20

* 0.54 excluding extreme points.

1.5 Summary of Findings and Recommendations

This report provides a wide range of findings and recommendations aimed at improving custom program performance and supporting CPUC and IOU program and policy enhancements for this important element of the IOUs' energy efficiency portfolios. Findings and recommendations were developed from each of the primary analysis activities. Extensive overarching findings and recommendations are presented in Chapter 8 of this report, while Chapters 5, 6, and 7 and Appendices C, D, and E, provide supplemental and more disaggregated suggestions (e.g., by IOU program group and measure group). At a summary level, the detailed recommendations in this report fall into the following primary areas:

- To address overestimation, on average, of ex ante savings estimates, the IOUs should:
  - Improve implementation and quality control of project operating conditions, ex-ante baseline determinations, calculation methods, and eligibility rules to address the discrepancy factors presented in this report, and
  - Improve adjustments to project savings based on post-installation inspections and M&V, where appropriate, while considering overall cost effectiveness constraints.
- To achieve sufficient quality control, increase consistency between project files and tracking data, and minimize miscommunication of project claims. IOU project documentation and tracking data needs to significantly improve.
- To reduce continued moderate to high free ridership, IOUs should design, implement, and test program features and procedural changes focused on increasing program-induced savings.

Finally, key recommendations discussed in Chapter 8 of this report are listed in Table 1-7.

Table 1-7: Summary of Key Recommendations

Key Recommendations by Topic Area
Operating Conditions
Increase focus on: a) accuracy of operating conditions, b) use of pre- and post-installation data and information, and c) keeping project documentation and tracking claims up to date with field information
Follow CPUC guidance to use observed conditions for savings calculations
Baseline Conditions
Increase efforts to ensure conformance with CPUC baseline policies
Eliminate common practice of defaulting to existing equipment baseline
Clearly identify project event in terms of natural replacement, replace on burnout, early replacement, new construction, etc., and set the appropriate baseline accordingly
Calculation Methods
Continue to review and improve impact methods and models through review of evaluation results, industry best practices, and collaboration with the CPUC's en ante review process
Carefully review ex-ante savings claims, inputs, and calculation methods
Identify fuel switching projects in claims and carry out the three-prong test. Identify and calculate impacts on all affected fuels. Improve QC on cogeneration-related projects and offer incentives only for grid savings in conformance with CPUC rules.
The IOUs should work with CPUC to define and promulgate appropriate impact estimation approaches with respect to whole building modeling and new construction
Cross-Cutting and Other Gross Impact-Related
Increase attention on compressed air, HVAC, and HVAC control projects
Review program-level GRR and LRA results and compare to internal IOU program performance assessments
Conduct periodic due diligence to ensure programs adhere to IOU and CPUC impact estimation policies, guidelines, and best practices. Continue to work closely and collaboratively with the CPUC's Ex Ante Review Process
Improve IOU program requirements, manuals, training, and quality control procedures in order to screen out ineligible projects
Extend M&V period and increase use of appropriate engineering analyses for MBCx

Key Recommendations by Topic Area
Documentation, Data Request, and Site Access-Related
Ensure consistency between tracking data and project files
Develop a final "closeout" report for each claim
Clearly identify all measures in project documentation files and tracking systems
Ensure data request responses are responsive, organized, clear and internally consistent
As a general guideline, all project documentation should be compiled in one electronic location
Prepare in advance for data requests to increase timeliness of response
Set up standard templates for project documentation and enforce compliance with these templates
Provide the final version of energy models and clearly identify the version of any simulation tool
Strengthen language in customer agreements making it mandatory for the customer, if selected, to participate in the CPUC's evaluation activities
Net-to-Gross/Program Influence
Adopt procedures to identify and affect projects with low program influence
Adjust the set of technologies that are eligible for incentives
Adopt procedures to limit or exclude known free riders
Make changes to the incentive design. Set Incentive Levels to Maximize Net (Not Gross) Program Impacts
Use a sophisticated program design reflecting a comprehensive mix of program features and leveraging an array of delivery channels
More information is needed on industrial project costs, non-energy costs and benefits, net present value analysis, and associated participant cost-effectiveness analysis

2

Introduction and Background

This report presents final[1] results from the impact evaluation of the 2010-2012 California IOUs' custom energy efficiency projects. This effort was managed by the CPUC Energy Division (ED) staff and is referenced as Itron Work Order Number 33 (WO033) on the CPUC ED public documents website.[2] The *Custom Impact WO033 Evaluation Plan*[3] was finalized on December 27, 2011 and, along with the five companion evaluation plan addenda, provides additional detail on the evaluation effort conducted; this evaluation plan and addenda are available on the ED public documents website. Readers may also want to familiarize themselves with two other relevant CPUC sources: a gross impact evaluation guidance document for custom and calculated projects[4] and a nonresidential Net-to-Gross (NTG) methods document,[5] which can also be found on the ED public documents website.[6] The scope of work for the evaluation of custom measures under WO033 includes an ex-post estimation of gross and net savings along with associated findings and recommendations that can be used to improve project and program performance and measure effectiveness.

This chapter provides background information and introduces the reader to the types of programs, facilities, and interventions evaluated under WO033. This chapter also references the research plan and evaluation architecture at a very high level. In the following subsections, we provide additional study background, highlight the percentage of portfolio claimed savings associated with this WO033 evaluation effort, and present the study objectives and issues researched.

---

1 This final report is comprehensive and supersedes an earlier interim reporting effort that was completed in January 2012. The interim report is located at: http://www.energydataweb.com/cpucFiles/pdaDocs/901/WO033_Interim_Report_Final%2001%2016%2013.docx

2 http://www.energydataweb.com/cpuc/home.aspx

3 http://www.energydataweb.com/cpucFiles/pdaDocs/814/WO33%20Research%20Plan%20Final%2012%2029.pdf

4 http://www.energydataweb.com/cpucFiles/pdaDocs/932/Evaluation%20Guidelines%20FINAL%202012.docx

5 http://www.energydataweb.com/cpucFiles/pdaDocs/910/Nonresidential%20NTGR%20Methods%202010-12%20101612.docx

6 The evaluation guidance and NTG methods documents were distributed and discussed with IOU project coordination group (PCG) and evaluation staff during the course of this evaluation, starting in 2011.

2.1 Background

This impact evaluation focuses on high priority evaluation objectives for custom programs and projects, including independent estimation of gross and net savings, provision of recommendations for program improvement, and reporting of ex-post results for use in CPUC cost effectiveness analyses.

More than 100 of the PY2010-2012 utility programs include custom, non-deemed projects. Some programs, such as the IOU commercial, industrial and agricultural calculated programs focus on custom or "calculated" incentives, while others provide a combination of deemed and calculated incentives. This evaluation effort investigates those custom measures and offerings across all IOU programs, including those undertaken by third parties or through local government partnerships. Although one priority objective of the evaluation is to estimate IOU fuel domain realization rates for custom projects across programs, the evaluation data collection and assessment efforts also support observations about specific programs where adequate sample sizes were obtained. This is especially true, by design, for the NTG sample and the lower rigor assessment sample (as outlined in Chapters 3 and 4).

The custom impact evaluation was organized into two periods — the first period covering calendar years 2010 and 2011, and the second period covering 2012 — to address the effect on custom program implementation of the CPUC's ex-ante-review-related Decision (D. 11-07-030).[7] The intent of this design[8] was to segment results in order to examine the effects of the ex-ante review (EAR) process and other custom project-related elements of D. 11-07-030 on project-level gross and net savings;[9] that is, whether any changes in ex ante project savings estimates due to the EAR process could be observed in evaluation results between the two periods. In addition, this design helped to ensure a reasonable representation of projects from both periods to create an overall sample representing PY2010-2012 cycle custom programs. This examination included an assessment of the IOUs' use of appropriate inputs and methods, as

---

7 http://docs.cpuc.ca.gov/published/FINAL_DECISION/139858.htm. Decision 11-07-030

8 The initial design sought to differentiate the 'before-decision' (BD) and 'after-decision' (AD) periods, with the BD period defined as all of 2010 and Q1 / Q2 2011, and the AD period defined as Q3 / Q4 2011 and all of 2012. However, this was later adjusted by calendar year to reflect the actual start of the full execution of EAR activities in January 2012.

9 The EAR process involves an M&V-level of review for IOU projects that are under development, prior to claims. CPUC staff and their contractors participate in these reviews and seek to actively influence the outcome of associated ex-ante project savings estimates, as well as IOU within-program engineering processes and procedures more generally. Importantly, D. 11-07-030 features detailed baseline requirements that were hypothesized to have significant influence on IOU project results, including remaining useful life/effective useful life (RUL/EUL) treatment and the need to demonstrate and document all associated early replacement (ER) claims.

well as an assessment of project-level execution within the confines of CPUC guidance and decisions and program rules.

The programs included in this custom impact evaluation address industrial and manufacturing facilities; water supply, water treatment and wastewater treatment; oil and gas extraction and refining; and commercial custom, non-deemed program offerings (including the Savings by Design new construction program and retro-commissioning programs). The scope addresses nonresidential custom measures of all types with one exception: lighting measures were generally excluded, except where a project's scope addressed the whole building (e.g., commercial new construction projects).

Each custom-oriented program offers one or more of the following interventions in order to encourage end users to upgrade to energy efficient measures: site specific facility assessments, feasibility studies, project incentives, facility audits, pump testing, and/or specialized training. For a more detailed description of the custom programs or measures addressed, please refer to the *Custom Impact WO033 Evaluation Plan*.

The CPUC organized all of its consultant evaluation and research work for PY2010-2012 into work orders, each of which was assigned a number. Many of these work orders addressed specific measures, sectors, or programs, while others addressed broader research topics such as baseline and market characterization research activities. To organize and define the impact evaluation related work orders, all measures in each IOU's portfolio were mapped to a measure group. Measure groups were then mapped and assigned to different work orders, each of which had its own project team, scope, and reporting. Work Order 33 was assigned all of the nonresidential custom projects, excluding lighting and codes and standards claims as mentioned above. Some deemed non-lighting, non-HVAC measures were also mapped and assigned to WO033 at the outset of the evaluation effort. These deemed measures were thought to fit best with the WO033 population since a given measure is often claimed using both deemed and non-deemed delivery channels. Examples of these deemed measures are refrigeration measures in grocery stores (common in the electric subpopulation) and steam traps (common in the gas subpopulation).

Energy savings claims from the measures assigned to WO033 represent a significant contribution to the overall savings portfolios for the IOUs' energy efficiency programs, accounting for about 24 percent of statewide electric savings claims and 72 percent of statewide gas savings claims during 2010-12. During this period, the IOU tracking data for measures assigned to WO033 included thousands of entries statewide with annual electric savings claims by the IOUs totaling 2,246 GWh and nearly 370 MW. Statewide IOU annual gas savings claims for measures assigned to WO033 total 148 million therms.

Table 2-1 reports claimed energy savings included in the custom impact evaluation using the final measure group mapping[10] to WO033 as discussed in the *Custom Impact WO033 Evaluation Plan*. The total claimed energy savings in Table 2-1 assigned to WO033, including deemed measures assigned to but not evaluated under WO033, are shown in Table 2-1.

Table 2-1: Claimed Energy Impacts by IOU for 2010-12 Projects in the Custom Impact Work Order (WO033)

Claimed Impacts by IOU			
IOU	**Electric Energy (GWh)**	**Electric Demand (MW)**	**Gas Energy (Million Therms)[11]**
WO033 2010-12 IOU Savings Claims			
PG&E	1,188	191	84
SCE	843	131	2
SDG&E	215	30	6
SCG	-	17	56
Total	2,246	368	148
Total 2010-12 IOU Savings Claims			
PG&E	4,053	722	108
SCE	4,500	831	5
SDG&E	791	130	10
SCG	21	27	83
Total	9,366	1,710	205
WO033 Percentage of Total IOU Savings Claims			
PG&E	29%	26%	77%
SCE	19%	16%	45%
SDG&E	27%	23%	64%
SCG	0%	61%	68%
Total	24%	22%	72%

Figure 2-1 presents the fraction of the WO033 ex-ante savings claims that were included in the evaluation. A fraction of the savings claims were removed from the evaluation scope as part of a data reduction step that is explained in greater detail in Chapter 3. For example, as a result of the CPUC's overall impact evaluation prioritization efforts the deemed non-lighting, non-HVAC

[10] CPUC consultants and staff worked together to create measure groups to facilitate the aggregation of like measures for the purposes of dividing the evaluation responsibilities by work order and to enable evaluation reporting by measure, where feasible.

[11] Gas savings reported includes only tracking records with positive therm impacts. A significant number of negative records in the complete sample frame are associated with increased heating due to the interactive effects of lighting efficiency measures. These records were not included in the sample population and sample design for gas projects, as their inclusion would otherwise allow a sample that would not accurately represent the actual number and aggregate savings of the natural gas projects implemented.

measures originally mapped to custom programs were not included in the ex-post impact evaluation activities of WO033. Only positive gas claims are shown.

The WO033 project population consists of over 50 measure groups and over 100 programs. As further explained in subsequent chapters, this presented challenges for allocation of evaluation resources and sample design with respect to different possible domains of analyses, e.g., IOU, fuel, custom versus deemed, program, or measure level. The portion of claims that were ultimately addressed in the WO033 impact evaluation activities is discussed in further detail in Chapter 3 and briefly introduced in Figure 2-1 and Figure 2-2.[12]

Figure 2-1: Ex-Ante Saving Claims Assigned to WO033

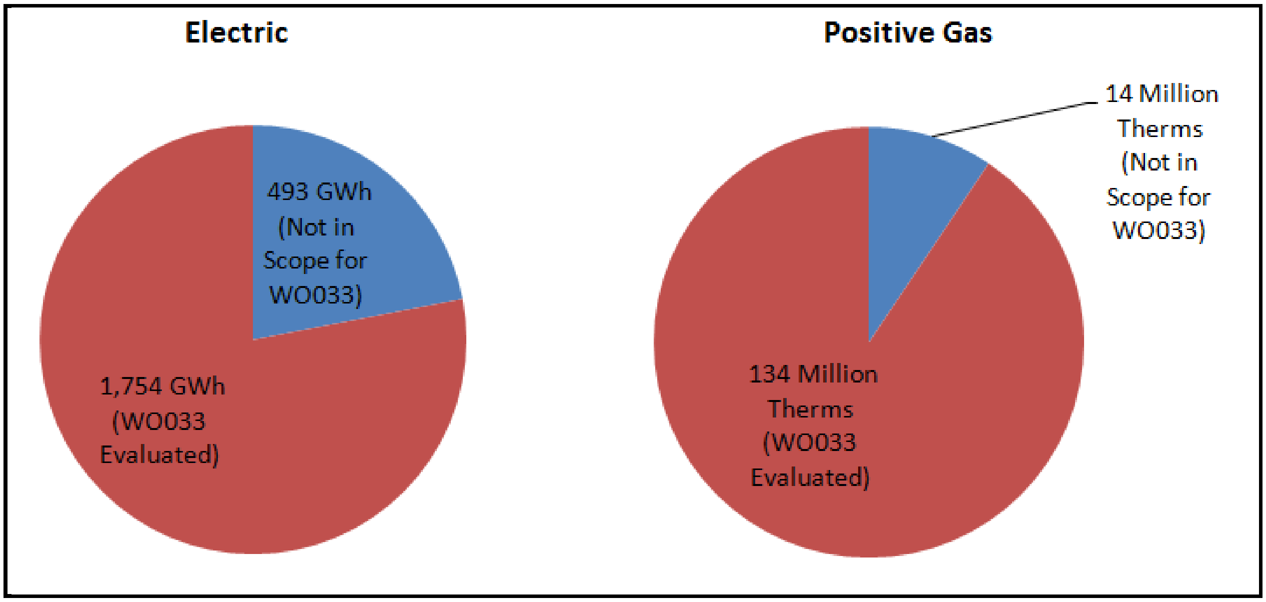
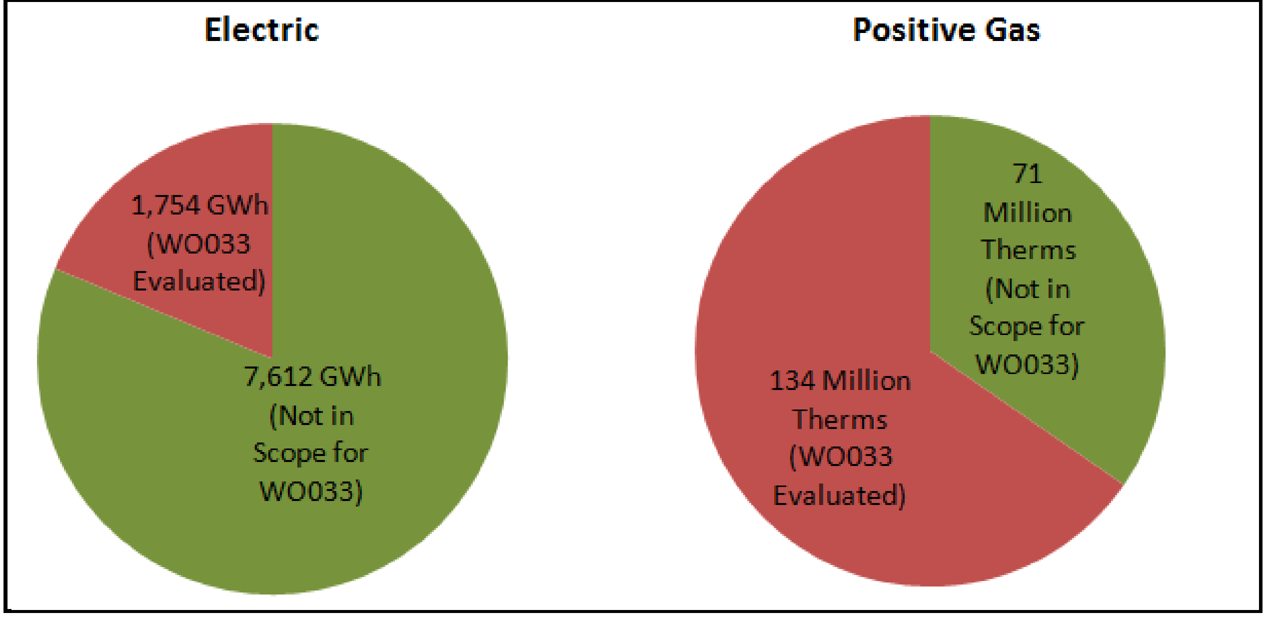


Figure 2-2 contrasts the WO033 ex-ante savings claims that were included in the evaluation with total portfolio claims for the 2010-2012 period. The savings claims included in the WO033 evaluation account for 19 percent of electric energy savings and 65 percent of positive gas claims.

[12] Please note that Table 2-1, Table 3-1, Figure 2-1, and Figure 2-2 reflect adjusted gross savings, i.e., savings claims adjusted by the 90% default GRR shown in Table 1-4. All other savings claims throughout the report and appendices reflect gross (unadjusted) savings claims, unless otherwise noted.

Figure 2-2: WO033 Evaluated Ex-Ante Savings Relative to Portfolio Claims



2.2 Study Objectives and Researchable Issues

The overarching goals and objectives of this impact evaluation are: to verify and validate the energy efficiency savings claims reported from IOU energy efficiency programs; to provide feedback on how well program procedures and savings calculation methods align with the CPUC's energy efficiency policies, requirements, and expectations; and to provide recommendations on how custom programs can be improved or refined. Gross energy savings, free ridership levels, and net energy savings (in kWh, kW and therms) are estimated and compared to IOU savings claims using evaluation-based realization rates and NTG ratios.

More details on the evaluation priorities[13] and the researchable issues for this effort are contained in the *Custom Impact WO033 Evaluation Plan*.

The priorities for this evaluation effort and the researchable issues include the following:

1. Estimate ex-post gross energy savings, characterize the factors that lead to the ex-post gross realization rates, and recommend how realization rates can be improved.
2. Estimate the level of free ridership, determine the factors that characterize free ridership, and provide recommendations on how program influence can be increased and/or free ridership reduced.

[13] These priorities include energy savings, net to gross ratios and program assessments.

3. Estimate participant spillover from efficient equipment installations made as a result of the program but without the provision of an incentive.
4. Provide feedback to the IOUs to improve program effectiveness.
5. Analyze the extent to which recommendations from the 2006-2008 evaluation cycle impact evaluations have been implemented by the IOUs and make further recommendations to close any gaps.
6. Determine whether the impact estimation methods, inputs, and procedures used by the IOUs and implementers are consistent with the CPUC's policy, guidance and decisions and best practices.[14]
7. Examine conformance with program rules, measure eligibility requirements, proper baseline specification, and other key project parameters/requirements (for example, data collection and reporting supporting dual baseline and early replacement claims).
8. Examine the extent of any program-induced acceleration of replacement of existing equipment and, in such cases, the RUL of the pre-existing equipment.
9. Collect data and develop information to assist with other research or study areas, including measure cost estimation, cost effectiveness analysis, strategic planning, and future program planning.

In order to more fully answer these researchable questions, this evaluation effort used a combination of approaches that included an expanded gross impact sample supplemented with desk review-only points (the latter desk reviews are part of what is referred to as the *lower rigor assessment* effort). Thus, in addition to the M&V-oriented gross impact work described in Chapters 3, 4, and 5, additional project level evaluation efforts included site-specific lower rigor assessment efforts. Lower rigor assessment methods and sample points are described in greater detail in Chapters 3, 4, and 7. These points were designed to expand the reach of the evaluation, through targeted sampling, to programs (and program groups) that would not otherwise receive much attention based on the size of the gross impact sample allocation alone. Both gross impact and 'lower rigor only' points were analyzed to: examine and comment on conformance with program procedures (including measure eligibility and other rules); analyze strengths and weaknesses of project applications; and provide feedback on impact estimation processes. For example, the project reviews and lower rigor assessments (LRA) examined whether: baseline specifications were appropriate; early replacement procedures were being adhered to; program eligibility rules were consistent with overarching regulatory guidance and program rules; and calculation methods and inputs were appropriate. This impact-related, program-level feedback also supported broader program assessments completed under Work Order 12 (Nonresidential

[14] See NR-5 Nonresidential Best Practices Report at www.eebestpractices.com

Process Evaluation) in this evaluation cycle. These results were featured in a WO012 report addressing calculated programs, the WO012 IOU Core Calculated Program Group Report.[15]

To more deeply explore the net impact effects and associated researchable questions noted above, the NTG sample frame and number of sample points was also strategically expanded as compared to previous nonresidential custom impact evaluation studies. Similar to the lower rigor efforts described above, NTG efforts included over-sampling of some medium sized programs, or groups of like programs, to support NTG ratio reporting at a program or program type level. Sample sizes designed for NTG ratio reporting were much greater than those for the gross impact and lower rigor activities. In so doing, the evaluation was able to support NTG results for a greater number of program segments than would ordinarily be available if standard sampling techniques were used for the population of energy efficiency projects assigned to WO033.

15 http://www.energydataweb.com/cpucFiles/pdaDocs/963/Non_Res_Core_Calculated_Prog_Assess.pdf

2.3 Structure of the Report

Table 2-2 shows the overall organizational structure of this report. Although findings and recommendations are overarching in Chapter 8, it is noteworthy that findings and recommendations are also included in Chapters 5, 6 and 7, as well as some of the appendices, including Appendices C, D and E. Readers seeking a more comprehensive assessment of opportunities for program improvement are therefore encouraged to read these particular chapters and appendices.

Table 2-2: Overall Organizational Structure of Report

Section #	Title	Content
1	**Executive Summary**	Summary of results and high level findings
2	**Introduction and Background**	Evaluation objectives, research issues, and savings claims
3	**Sample Design**	Sampling design and associated issues
4	**Methods**	Approaches to gross impact determination, on-site M&V activities, NTG surveys, and lower rigor assessment (desk review) activities
5	**Goss Impact Results**	Gross impacts and realization rates, measure and program differentiation, new construction highlights
6	**Net Impact Results**	Net of free ridership ratios and results, spillover results, net realization rates and NTG result drivers
7	**LRA Results**	Program assessments based on project documentation review using the lower rigor approach
8	**Detailed Findings and Recommendations**	Presented by topic area, including operating conditions, baseline issues, calculation methods, cross-cutting, net-to-gross/program influence, and lower rigor program related,

3

Sample Design

In this chapter, we present an overview and summary of the sample design used for the Custom Impact Evaluation. The sample design for this study was updated several times during the course of the evaluation; a more detailed explanation for these adjustments is included in Appendix A. This chapter presents the final design. These sampling changes and other key research plan modifications are discussed in greater detail in Appendix A, Research Planning and Modification. More detailed information on the allocation of net-to-gross (NTG) and lower rigor (LR) samples is provided in Chapters 6 and 7, respectively, and further supported by information in Appendices D and E.

A number of study objectives and considerations informed the sample design and associated data collection effort. Two key aspects were consideration of project and resource constraints and the number of domains for which statistically reliable results could be produced. For example, implementation of a sample design that would optimize the likelihood that the results will meet desired overall statistical confidence and precision levels, while simultaneously providing meaningful information and results for selected programs or groups of programs, would likely place constraints on the level of effort and the amount of data that could be collected and analyzed, given the number of custom-related programs in 2010-2012. Examples of other sample design-oriented considerations included the expected error ratio (or coefficient of variation) of the sample, the number of sampling variables and domains, the finite population size of the domains, the cost per point of the data collection, the expected accuracy of ex post results, and the total budget available for the study. Other important questions that were considered included the following:

- How should different policy and program objectives for custom programs and measures be addressed?
- How might different sampling activities be constructed to address different evaluation needs and cost-related constraints?
- How should different program years or policy periods be considered?

3.1 Custom Sample Frame Data Reduction

Due to the extensive and heterogeneous nature of the WO033 population of projects, it was recognized that tradeoffs and priorities needed to be established to better focus the evaluation effort. For the purposes of gross impact, lower rigor, and NTG data collection, certain assigned records were removed from the WO033 sample frame. This included removal of the following project records from the population for the ex post impact evaluation activities for this work order:

- Deemed measures
- Pump testing records
- Stand-alone progress payments for NRNC projects[1]
- Tracking records that appeared in the final IOU upload that were not included as tracking records in previous IOU uploads used for sampling

The extent of this data reduction effort is presented in Table 3-1. The remaining records constitute the custom impact sample frame from which projects were selected for primary data collection and related evaluation activities.

The sampling unit used for the sample implementation is also introduced in this table. Each sampling unit is a combination of one or more tracking system records that together best represent a *project*. A project, for example, might reflect all related activities at a site for a specific measure or a fundamentally integrated group of measures, such as the individual measure components of a central plant upgrade or renovation. In general, sampling units reflect a unique measure name or application number associated with a specific site ID[2] Further refinements were applied using judgment through the review of each record in the population. Despite the use of coding algorithms, manual review was necessary to ensure that all WO033-assigned measures related to a given project were included in each sampling unit. This was necessary due to inconsistencies in how tracking records are defined across utilities and programs. Sample points were selected for evaluation using the resultant sampling units after this tracking system consolidation step.

1 These tracking records did not include impact claims but did include incentive payments and were not tied to other NRNC tracking records with claimed savings at this time. These appeared to be projects in progress.

2 Applications consisting of more than one tracking record for a given site ID were defined as a sampling unit. Applications by site ID were only grouped to form a sampling unit for applications that consisted of a single record and where applications share the same IOU measure name. However, applications were generally not grouped for Monitoring-based Building Commissioning (MBCx) and whole building new construction records, as records typically consisted of unique buildings, and buildings were chosen as the appropriate sampling unit.

Table 3-1: Tracking System Data Reduction Steps

Population	Data Reduction	Utility	2010-12 Tracking Records	Sampling Units***	Positive Electric Energy Savings (MWh)	Positive Gas Energy Savings (Thousands of Therms)	Total Incentive (Thousand $)
WO033*	NA	All	134,300	NA	2,246,389	147,947	393,563
		PG&E	65,617		1,188,121	83,555	206,821
		SCE	43,985		839,459	2,174	112,276
		SCG	2,056		0	55,898	32,820
		SDG&E	22,642		218,809	6,320	41,646
Records Removed							
WO033 Subpopulations	Deemed measures	PG&E	53,535	NA	275,859	7,892	36,971
	Deemed measures	SCE	25,949		146,907	12	31,696
	Deemed measures	SCG	1,001		0	5,064	2,252
	Deemed measures	SDG&E	18,829		31,782	807	8,461
	Stand-alone NRNC progress payments	PG&E	432		0	0	3,351
	Stand-alone NRNC progress payments	SCE	10		0	0	1,076
	Stand-alone NRNC progress payments	SCG	0		0	0	0
	Stand-alone NRNC progress payments	SDG&E	0		0	0	0
	SDGE3170 Records****	SDG&E	53		10,096	131	1,251
	Pump testing	PG&E	721		0	0	153
	Pump testing	SCE	14,066		28,146	0	0
Remaining Custom Records in Sample Frame							
WO033 Custom**	NA	All	19,704	12,778	1,753,599	134,041	308,352
		PG&E	10,929	7,448	912,261	75,663	166,346
		SCE	3,960	3,063	664,406	2,162	79,503
		SCG	1,055	731	0	50,834	30,569
		SDG&E	3,760	1,536	176,931	5,382	31,934

* Represents all tracking records assigned to WO033 using ED measure group assignments, as discussed in Chapter 2.

** Represents all tracking records that remain in the custom WO033 sample frame.

*** Sampling unit is generally a unique measure name or application number by site id. Further refinements are applied using judgment.

**** These records were not present in the March 2013 extract, which was used for sampling. They were added to the Final extract dated July 2013, when sampling had been completed.

3.2 Overall Sampling Plan

Sample designs were developed for the gross impact, LR, and the NTG elements of the evaluation. Targeted sample sizes for both the BD and AD periods and each type of sample point are shown in Table 3-2 below. A discussion of the considerations for development of these samples is provided in the following subsections, as well as the chapters and appendices mentioned earlier in this chapter. Actual samples realized are discussed in each of the evaluation results chapters.

Table 3-2: Summary of Overall Sample Sizes for WO033 Impact-Related Effort

Impact Evaluation Component	Before-Decision	After-Decision	TOTAL
M&V Points (Gross Realization-Rates + NTG)	200	200	400
Overlapping M&V Points (GRR + NTG)*	0	50	50
Lower Rigor Points (Qualitative + NTG)	100	0	100
Incremental NTG-Only Points	480	350	830
TOTAL**	**780**	**600**	**1,380**

* M&V points that were selected for gross impact efforts and were previously selected for ex-ante review or early opinion review under WO002 are designated 'overlapping' points.

** All points incorporate NTG evaluation in addition to gross impact evaluation efforts.

All gross impact points included lower rigor evaluation, such that the total lower rigor sample is equal to gross impact points plus lower rigor points, for a total target sample size of 550 points. All gross impact and lower rigor points also included NTG data collection, such that the total NTG sample is equal to gross impact plus lower rigor plus incremental NTG-only points, for a total target sample size of 1,380 points.

For points that are claimed for savings under WO033 and that were formerly *Ex-Ante Review Points*[3] or *Early Opinion Requests*,[4] a census was undertaken for the ex-post gross impact effort. These points are directly affected by the WO002 review process and are denoted as overlapping points, with a targeted number of completed overlapping points originally estimated at 50 points.

---

[3] These WO002 assessments entail the review of the IOU and third party custom measure documentation as a project progresses from the initial application pipeline, i.e., project development, to its completion and final savings claim. The EAR process "freezes" the final ED-reviewed and ED-approved project ex-ante savings, and thereby effects the final IOU incentive payments, since custom incentives are paid per unit of energy saved.

[4] These WO002 assessments entail an IOU-requested review by the EAR team of the IOU project documents from the initial pipeline applications. The early opinion assessments provide guidance that could affect the IOU ex-ante savings estimates, but freezing of savings is not always undertaken.

3.3 Periodic Sampling

The WO033 Custom Impact evaluation was conducted in parallel with program execution by the IOUs. As the IOUs made progress and reported claims associated with tracking system records, the evaluation completed sampling for gross impact and NTG on a periodic basis for selected periods of program performance. There were four sampling periods: BD (Q1 2010 – Q2 2011), AD1 (Q3/4 2011), AD2 (Q1/2 2012), and AD3 (Q3/4 2012).[5] Supplemental lower rigor points were only sampled from the first period, Q1 2010 – Q2 2011. Details regarding the specifics of M&V, lower rigor, and NTG sampling are explained in greater detail below.

3.4 Gross Impact Sampling

The sampling effort that was most constrained was the gross impact sample, as this sample had a relatively high cost per point due to the requirements associated with estimating savings for large and complex custom projects in the nonresidential sector. This relatively high cost per gross impact point was required to meet the objective of reducing measurement error (which requires extensive project review, on-site visits, and, to varying degrees, short and mid-term measurement). Given the expected error ratio (between 0.5 and 1.0) and the desire for a relative precision of 10 percent at the 90 percent confidence interval for the combined BD and AD periods, the number of potential sampling domains was limited. Given possible sampling domains considered,[6] IOU and fuel type were ultimately chosen for GRR results, with the sample segmented into two periods, 200 points for BD and 200 points for AD.[7] This approach resulted in five sampling domains for which gross realization rate reporting was targeted, i.e., PG&E electric, PG&E gas, SCE electric, SCG/SDG&E gas, and SDG&E electric.[8]

---

5 Sampling periods were developed to allow impact evaluation activities to start as quickly as possible following the evaluation planning phase and then to sample every six months. There were several benefits to this design. The first period – the BD period –reflected claims prior to Decision D.11-07-030. The segmented sampling periods allowed for ongoing data collection efforts and the ability to spread out the work and thereby complete a greater volume of project work in a more timely fashion. The alternative of waiting for all participation to be completed prior to sampling was not considered a reasonable approach and would have greatly delayed project completion for a similar size effort. However, there were some challenges and costs to this approach: participants were sometimes selected on multiple occasions on the gross and net impact evaluations, in efforts that could possibly have been combined into one project or sampling unit; in addition, waiting for final claims would have helped to streamline sampling and administrative efforts due to a number of factors, including IOU true up of savings claims that were made.

6 A wide range of potential sampling domains were considered for the gross impact (gross realization rate) sample, including: IOU, fuel (electric, gas), energy metric (e.g., kWh, kW), measure group or HIM, end use, sector, program, period (e.g., year, before-/after- CPUC ex-ante review related Decision D.11-07-030), program or measure goals, and future potential, as well as combinations among these.

7 In addition, 50 gross impact points were also allocated to projects that had been selected for and evaluated in the ex-ante review process and that were subsequently installed during 2010-12.

8 Although desirable, it was not feasible to design the gross impact sample around measures or programs with an adequate level of statistical accuracy.

The first sample selection was conducted for the BD period. The sampling units in each IOU-fuel domain were sorted in descending order of kWh or therms ex ante savings, and were placed in five strata, where stratum 1 included the largest projects and stratum 5 included the smallest projects. The strata boundaries were defined in such a way that, for the BD period, each stratum represented approximately 20 percent of the total population savings. For each IOU-fuel domain, sample points were then randomly selected in each stratum. For consistency across sampling periods, the strata boundaries selected in the BD period were preserved for sampling purposes during subsequent sampling periods. Strata boundaries are shown in Table 3-3.

Table 3-3: Strata Boundaries Defined for WO033 Sampling Efforts

	Electric						Gas			
	PG&E		SCE		SDGE		PG&E		SDGE	
	Strata Boundaries (kWh)						Strata Boundaries (Therms)			
Strata	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
1	3,542,350	∞	3,821,945	∞	1,837,042	∞	10,205,170	∞	4,790,381	∞
2	1,355,232	3,542,350	2,062,485	3,821,945	645,647	1,837,042	3,321,543	10,205,170	708,450	4,790,381
3	580,266	1,355,232	737,621	2,062,485	394,404	645,647	1,040,884	3,321,543	212,440	708,450
4	186,610	580,266	246,026	737,621	134,007	394,404	196,030	1,040,884	66,113	212,440
5	0	186,610	0	246,026	0	134,007	0	196,030	0	66,113

3.5 Sampling and Extrapolation Methods, Confidence and Precision

The ratio-estimation approach for sample design described in Chapter 13 of the *Evaluation Framework Study*[9] and referenced in the *California Energy Efficiency Evaluation Protocols*[10] was selected to develop program realization rates in the current evaluation cycle. A key input to the ratio-estimation sample planning methodology is the error ratio (*er*) that is expected to result given the evaluation sample size selected (the *error ratio* is defined below). As with the a priori use of the expected coefficient of variation in other sampling methods, the variance in the parameter of interest is not known prior to completing the evaluation work. Instead, analysts must estimate the *er* from other related studies and work or summarize expected sampling results across a range of possible *er* (as is often done with confidence levels).

To more formally investigate the expected precision levels for the current evaluation the precision level achieved for the 2006-2008 PG&E Fab impact evaluation sample was first reviewed. The precision estimation process was carried out as described for ratio estimation-

9 http://www.calmac.org/publications/ California_Evaluation_Framework_June_2004.pdf

10 Chapter 13 – Sampling, page 358, of the TecMarket Works, 2004. *2002 Evaluation Framework Study*, prepared by TecMarket Works for Southern California Edison Company, June.

based samples in Chapter 13 of the *Evaluation Framework Study*. Specifically, the error ratio was calculated and the precision expected was estimated, with alternative sample sizes as described on pages 358 and 365 of the *Study*, respectively, using the results from the 2006-2008 PG&E Fab impact evaluation.[11] From this study an error ratio (*er*) was calculated using the following formula:

$$\hat{er} = \frac{\sqrt{\left(\sum_{i=1}^{n} w_i \ e_i^2 / x_i^{\gamma}\right)\left(\sum_{i=1}^{n} w_i \ x_i^{\gamma}\right)}}{\sum_{i=1}^{n} w_i \ y_i}$$

where

$\gamma = 0.8$

$e_i = y_i - \hat{B} \ x_i$

w_i is the case weight,

x is the tracking estimate of savings for each project, and

y is an estimate of the estimated savings from the ex post evaluation.

Based again on the 2006-2008 work, case weights were used to calculate the stratified ratio estimator of B, denoted $\widehat{B}$, as follows:

$$\hat{B} = \frac{\hat{Y}}{\hat{X}} = \frac{\sum_{i=1}^{n} w_i \ y_i}{\sum_{i=1}^{n} w_i \ x_i}$$

Then the relative precision of $\hat{B}$, at the 95 and 90 percent confidence levels, can be calculated for alternative sample sizes using the equation below (which includes finite population correction):

[11] See Chapter 3 of Itron, 2010. *2006-2008 Evaluation Report for PG&E Fabrication, Process and Manufacturing Contract Group*, prepared by Itron, Inc. for the California Public Utilities Commission, February.

$$rp = 1.96\sqrt{1-\frac{n}{N}}\frac{er}{\sqrt{n}} \quad \text{95\% CL}$$

$$rp = 1.645\sqrt{1-\frac{n}{N}}\frac{er}{\sqrt{n}} \quad \text{90\% CL}$$

Based on the 2006-2008 evaluation results, the error ratio expected for the 2010-2012 impact sample and analysis was hypothesized to be in the order of 0.80. Because error ratio have varied across previously conducted studies and study domains, a range of error ratio was used in the initial research planning process to assess the effect of high and lower ratios. In the end, the 0.8 was selected as the point estimate used for the final sample design and budgeting process.

Table 3-3 and Table 3-4 present the sample design-based expected precision, population, and gross impact points allocated for the WO033 non-overlapping gross impact sample by sampling domain. Expected precision is shown for the combined BD and AD periods.

Table 3-3 features targeted sampling precision by IOU-electric domain, while Table 3-4 presents targeted sampling precision by IOU-gas domain. These tables present the sample design, including forecasted participation at the time of the design, not the resultant sample. Resultant sample statistics are presented in Chapter 5 and include the actual number of sampling units.

Table 3-4: WO033 Gross Impact Sampling Precision by IOU-Electric Domain

Utility	Positive Electric Sampling Units (N)*	Gross Impact Allocation (n)	Assumed Error Ratio**	Expected Precision at 90% C.I.	Expected Precision at 80% C.I.
PG&E	6,320	100	0.80	13%	10%
SCE	2,298	100	0.80	13%	10%
SCG	0	0	-	-	-
SDG&E	996	60	0.80	17%	13%
Total Electric	9,614	260	-	8%	6%

* Forecasted figure.

** Error ratio estimate based on results from the *2006-2008 PG&E Fabrication and Manufacturing Contract Group* results.

Table 3-5: WO033 Gross Impact Sampling Precision by IOU-Gas Domain

Utility	Positive Gas Sampling Units (N)*	Gross Impact Allocation (n)	Assumed Error Ratio**	Expected Precision at 90% C.I.	Expected Precision at 80% C.I.
PG&E	532	80	0.80	14%	11%
SCE	12	0	-	-	-
SCG / SDG&E	374	60	0.80	16%	12%
Total Gas	918	140	0.80	10%	8%

* Forecasted figure.

** Error ratio estimate based on results from the *2006-2008 PG&E Fabrication and Manufacturing Contract Group* results.

This same method was used to estimate and report the relative precision and error ratio for the samples selected in 2010-2012. As shown in Chapter 5, the resulting error ratios range from 0.56 (for the PGE Gas domain) to 4.44 (for SDGE/SCG Gas domain); however, when a small number of extreme realization rates are excluded, the error ratios for the primary design metric, energy, tended to be fairly close to the 0.8 value used for sample planning (0.56 to 1.03).

Once M&V was completed, individual project gross realization rates (GRRs) were aggregated to the IOU-fuel domain level using two types of weights:

- To estimate stratum-level GRRs each sampled project was attributed a **project weight**. For projects that were completely evaluated (M&V completed for all records of a given project,) the project weight was set equal to the sum of ex ante savings of the records. For projects that were only partially evaluated (for example: M&V completed for only one of five records in the project), the project weight was set equal to the sum of ex ante savings of the records that were evaluated.
- To estimate the GRR at IOU-fuel level, stratum-level GRRs were weighed using **strata weights**; strata weights were calculated as the sum of ex ante savings for all projects allocated to a given stratum.

3.6 Lower Rigor Analysis and NTG Sampling

Net-to-gross surveys and project-level net-to-gross analyses were targeted for all of the gross impact and lower rigor points. As a result of the relatively low marginal cost of additional telephone surveys, hundreds of additional NTG sample points were included beyond the gross impact and lower rigor points. Supplemental NTG points were directed to individual programs or groupings of programs. The resulting larger NTG sample enables reporting of NTG results for a larger number of programs/program groups than would be possible if the NTG sample mirrored the gross impact sample.

As discussed in Section 8 of the *Custom Impact WO033 Evaluation Plan*[12] referenced earlier in this report, in order for the evaluation to provide input for selected programs, a minimum of 10 to 15 points per program was targeted for LRA program samples and 25 to 50 points per program for NTG program samples across the program cycle (i.e., across the BD and AD periods). While the NTG samples sought to achieve targeted precision levels (e.g., 90/10 to 90/20), the LRA sample was intended more as a due diligence sample to monitor compliance with regulatory and program rules for savings estimation. As expected, actual allocations by program domain differed somewhat from these general targets. First, core and other larger programs sometimes met these targets based upon the random gross impact allocation alone. Moreover, certain programs with a relatively small number of participants were allocated a lower number of points due to the benefits of small populations on targeted precision levels.[13] Allocations to program and program groups for the NTG and LR efforts are described in Chapters 6 and 7, respectively and are further supported in Appendices D and E. Sample points were selected by size stratum within each program group (see below) for which extra NTG or LRA points were allocated. NTG results were weighted up using the same approach described above for the impact sample.

3.7 LRA Programs and Groups

The population of projects incorporated under WO033 is diverse and heterogeneous in nature, consisting of over 50 CPUC-defined measure groups[14] (ED measure groups) and tracking records from more than 100 individual programs. Our detailed evaluation planning activity involved an assessment of general sampling options, given the costs of various evaluation activities. As discussed above, it quickly became apparent that for the most expensive activity,

---

12 www.energydataweb.com/cpucFiles/pdaDocs/814/WO33%20Research%20Plan%20Final%2012%2029.pdf

13 When estimating sample sizes from small populations, or when sampling a large portion of the population, a finite population correction (FPC) factor is multiplied by the standard error, which reduces the estimated confidence interval. Where applicable using FPC, targeted precision levels are achieved using a smaller sample size. The FPC is calculated as the square root of the ratio of (the population minus the sample size) and (the population minus one).

14 It is important to also remind the reader that the custom impact Work Order 33 (WO033) population of tracking records is defined based on a mapping of tracking records to CPUC-defined ED measure groups and a subsequent mapping of CPUC measure groups to each work order. Measure groups were also segmented and allocated by residential (versus nonresidential) and deemed (versus non-deemed). See the *2010-2012 Energy Efficiency Evaluation, Measurement and Verification Work Plan* (available on www.energydataweb) for a discussion of ED measure group assignments. Assignment of tracking system records to the WO033 sample frame is discussed in the *Custom Impact WO033 BD Period Sampling Addendum 2010-2012 Impact Evaluation* available at the same location.

estimation of gross impact realization rates through site-specific M&V, it would be best to design the sample around a relatively small number of domains but support gross impact-related results at a more detailed program and measure level through additional targeted activities, namely, lower rigor analyses.

To complement the gross impact sample and gain further insight into program or measure performance, the decision was made to selectively sample certain programs (or program groups) of interest. However, due to the expensive nature of M&V, the evaluation team reduced the rigor of these supplemental points by essentially conducting a project practices assessment using an application or ‘desk’ review. As previously noted, these supplemental points were labeled lower rigor assessment (LRA) points. The lower rigor points involved desk reviews of IOU-provided application and program documents to assess ex-ante savings and baseline-related methods, procedures, estimates, and assumptions. The additional information garnered from lower rigor points is primarily qualitative in nature. These lower rigor points do not yield ex-post energy savings estimates or gross realization rate (GRR) estimates. To reach the sample size needed to adequately represent programs and program groups of interest and to provide for early feedback on program performance, the lower rigor effort involved all 200 BD gross impact points and 100 supplemental (lower rigor only) points randomly chosen but targeted at specific programs. In addition, for the final report, the AD gross impact points also contributed to the lower rigor assessment effort.

3.8 Program Group Domains

For both the LRA and NTG efforts, program domains were configured to be individual programs for a given utility, a group of like programs within a utility, or a statewide grouping of like programs across IOUs. The rationale for program domain selection includes the following considerations: input from the IOUs; interest from the CPUC; program energy saving claims and associated goals for the full cycle; the fraction of claims associated with measures included in WO033; core versus third party delivery mechanisms; a non-residential focus; and a custom or non-deemed emphasis. Measure types or measure groups were not primary or secondary sampling criteria.

Table 3-5 presents the selected programs and program domains, by IOU, for which results are available (including programs or domains where supplemental LR and NTG sampling was completed). Examples of programs that were not selected include programs that emphasize deemed savings (consistent with sample frame data reduction steps outlined previously),[15] agricultural pump testing, and programs that are very small. Some of the selected program

[15] Lighting and small HVAC deemed measures are mapped to WO029 and WO032, respectively. The remaining deemed measures in WO033 are primarily refrigeration measures for electricity and steam traps and pipe/tank insulation for gas.

domains represent groups of programs, such as the local government partnerships, and the third party programs; other groups include the CSU/UC programs and retro-commissioning programs. IOU core and some third party programs are maintained as separate program domains. Additional details on programs and program domains are in Chapters 6 and 7 and in Appendices D and E.

Table 3-6: WO033 Program Groups by IOU and Statewide

Program Names, Numbers and Aggregated Groupings
PGE21011 – Commercial Calculated Incentives
PGE21021 - Industrial Calculated Incentives
PGE21031 - Agricultural Calculated Incentives
PGE21042 - Savings By Design
PGE2222 - Energy Efficiency Services for Oil Production
PGE2225 - Refinery Energy Efficiency Program
PGE2223 - Heavy Industry Energy Efficiency Program
PGE Other Third Party (3P)
PGE RCx (Retro-commissioning)
SCE-SW-003B – Industrial Calculated Energy Efficiency Program
SCE-SW-002B – Commercial Calculated Energy Efficiency Program
SCE-SW-004B - Agriculture Calculated Energy Efficiency Program
SCE-SW-005A - Savings By Design (New Construction)
SCE UC/CSU Group
SCE Other Third Party (3P)
SCE Local Government (LG)
SCG 3625 - NRNC Savings By Design
SCG Core Calculated
SCG Third Party
SDGE3117 - Local03 - Local Non-Residential (BID)
SDGE3118 - SW-NCNR - NRNC Savings By Design
SDGE Core Calculated
Statewide University of California/California State University (SW UC/CSU)
Statewide Energy Watch/Local Government (SW EW/LG)
Statewide California Community College (SW CCC)
Statewide California Department of Corrections (SW CA DOC)
Statewide State of California (SW CA State)

4

Methods

Methods used in the evaluation are described in this chapter, including an examination of data sources and constraints. The discussion addresses the gross impact, net impact, and LRA efforts.

4.1 Data Sources and Constraints

This section briefly outlines primary and secondary sources of information that were examined in conducting this evaluation.

A variety of primary data sources were used in the gross and net impact calculations in order to develop robust, accurate, and defensible ex-post estimates of measure impacts. Data sources provided by the IOUs included tracking data, audit reports, incentive applications, IOU billing information, building energy simulation and other models used in the ex-ante analysis, and other project documentation, including field measurements to support the project savings estimates. These data sources were used for both the gross impact effort and the LRA effort. Data sources collected and documented by the evaluation team included telephone interviews, on-site interviews, visual inspection of the systems and equipment, collection of customer supplied data, EMS (Energy Management System) and SCADA (Supervisory Control and Data Acquisition) system data downloads, spot measurements, industry standard practice (ISP) investigations, short-term monitoring (for example, less than four weeks), and mid-term monitoring (for example, four to eight weeks).

For the NTG effort, three primary data sources were used to examine free-ridership and spillover:

1. **Program Files**. Custom programs maintain a paper or electronic file for each paid application. These files can contain various pieces of information that are relevant to the analysis of free-ridership such as letters written by the utility’s customer representatives that document what the customer had planned to do in the absence of the rebate or explain the customer's motivation for implementing the efficiency measure. Information on the measure payback with and without the rebate was also available in some of the files.

2. **Decision-Maker Surveys**. When a site is recruited, one must also determine who was involved in the decision-making process that led to the implementation of measures under

the program. They are asked to complete a decision maker (telephone) survey. This survey includes highly structured questions addressing the probability that the customer would have implemented the same measure in the absence of the program.

3. **Utility and Program Staff Interviews**. For the largest and most complex projects, additional interviews were conducted with program representatives and utility account representatives. These interviews were designed to gather information on the historical background of the customer's decision to install the efficient equipment, the role of the utility and program staff in this decision, and the name and contact information of vendors who were involved in the specification and installation of the equipment.

In addition, targeted interviews with market actors (such as equipment suppliers) were conducted to determine standard practice for particular projects where warranted to establish project baselines. Additional data sources included IOU program tracking data and IOU-provided project specific documentation.

Secondary data sources were also utilized, as described in more detail in Section 4 of the *WO033 Custom Impact Evaluation Plan*. Examples of secondary data used in this study include other evaluation studies, including the 2006-2008 custom impact evaluations for the CPUC; DEER and DEER updates; manufacturers' literature and equipment specifications; conference papers; and standard technical references from industry organizations such as ASHRAE and DOE.

Data constraints included: inadequate IOU measure descriptions; undocumented IOU baseline determinations; lack of remaining useful life indicators; and poor or missing IOU or customer contact information. In addition, poor, incomplete project documentation with regard to savings calculations or final savings claims were common. In selected cases, data constraints were attributable to participants and their contractors not being able to provide sufficient time for site access, not having knowledgeable personnel available for site visits, not allowing access for M&V data collection including logger installation, not having access to records, and / or not collecting or being able to trend EMS or SCADA data.

4.2 Overview of Impact Evaluation Methodology

This evaluation used approaches to estimating ex-post gross and net energy savings similar to those used in the 2006-2008 custom impact evaluations for the CPUC (in particular, the *2006-2008 Evaluation Report for PG&E Fabrication, Process and Manufacturing Contract Group),*[1] relying primarily on site-specific measurement and verification for the gross impact evaluation

[1] http://www.calmac.org/publications/PG&E_Fab_06-08_Eval_Final_Report.pdf

and decision maker surveys for the net impact evaluation. The key steps used to develop gross and net savings estimates for the custom program population were to:

- Independently verify reported measure installation for each project in the gross impact sample.
- Develop ex-post estimates of the energy savings for each project in the sample.
- Develop net impact estimates in the form of net-to-gross ratios in an expanded net-to-gross / net of free ridership impact sample.
- Apply those findings to the full participant population.

For the gross impact M&V effort at sampled participant sites, the engineering analysis methods and degree of monitoring varied, depending on the complexity of the measure, the size of the associated savings, and the availability and reliability of existing data. For net impact determination, the level of effort also varied by project size and complexity with the largest projects receiving an in-depth level of inquiry and small/medium-sized projects receiving a more basic examination. The lower rigor assessment effort was used to provide additional, qualitative reviews of programs and program groups at an IOU and statewide level.

The methodologies for the gross impact, net impact, and LRA efforts are described in the following sections.

4.3 Gross Energy and Demand Impact Evaluation Activities

The custom measure gross energy and demand impact evaluation involves a range of evaluation and M&V approaches, including on-site data collection, monitoring, and analysis for a representative sample of custom projects. For additional details on gross impact methods used, refer to the *Custom Impact WO033 Evaluation Plan*[2] and the *Procedures for Site-Specific Impact Analysis under Work Order WO033.*[3] The gross impact effort was guided by the following sources, which were referenced frequently or used comprehensively in this evaluation:

- *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*[4]
- *The California Evaluation Framework*[5]

[2] http://www.energydataweb.com/cpucFiles/pdaDocs/814/WO33%20Research%20Plan%20Final%2012%2029.pdf

[3] http://www.energydataweb.com/cpucFiles/pdaDocs/885/WO033%20Custom%20Impact%20Procedures_Jan2012_v1.docx

[4] http://calmac.org/publications/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006ES.pdf

[5] http://www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf

- The CPUC-ED gross impact evaluation guidance document[6]
- Dispositions of projects submitted for ex-ante review
- *Methodological Framework for Using the Self-Report Approach to Estimating Net-to-Gross Ratios for Nonresidential Customers*[7]

A distinct set of steps, as listed below, were employed to produce gross impact results. The evaluators:

- Obtained the application files and associated documentation from the IOUs for each gross impact sample point.
- Completed a project documentation review form that was the basis of the Low Rigor Assessment (LRA).
- Prepared a Site-Specific Measurement and Verification Plan (SSMVP) for internal review and input from CPUC staff and consultants.
- Scheduled and conducted on-site visits (including project verification, on-site interviews, data collection, monitoring, and so forth).
- Integrated the gross impact findings with the net to gross findings for each gross impact project to properly assign the effects of baseline, industry standard practice, and participating company standard practice.
- Developed ex-post impact estimates for each site.
- Prepared detailed, site-specific impact evaluation reports for internal review and input from the CPUC staff and consultants.
- Entered summary impact results and discrepancy factors into a "findings" workbook to facilitate cross-site analysis.
- Extrapolated the final ex-post realization rate estimates for the sample to the remaining applications in the sample frame population.

4.3.1 Data Request for IOU Project Documentation

The data request template for project documentation listed 21 items that were identified by the evaluation team as necessary to provide the full breadth of information for conducting evaluation activities. Those items included the paperwork originally submitted to request funding for the project, data archives collected before and after the project, building energy simulation or other

6 http://www.energydataweb.com/cpucFiles/pdaDocs/932/Evaluation%20Guidelines%20FINAL%202012.docx

7 http://www.energydataweb.com/cpucFiles/pdaDocs/910/Nonresidential%20NTGR%20Methods%202010-12%20101612.docx

technology-specific models that were used to develop ex-ante savings, and drawings or specifications. The specific list included in the cover letter of the initial project data request sent to each IOU is provided in Appendix C. Supplemental project data requests were presented to the IOUs for specific items that were not provided in the initial data request but were required for thorough, comprehensive, and accurate evaluation efforts for specific projects.

4.3.2 IOU Project Documentation Review

For each selected gross impact sample point, the assigned engineer performed an application review to assess the engineering methods, parameters, calculations, and assumptions used to generate the ex-ante impact estimates. This step also allowed for an assessment of additional data needs from the IOU and monitoring needs required to complete the analysis and ex-post savings estimation for each individual gross impact sample point. The form used for the individual project lower rigor assessment was completed for each gross impact sample point after the initial documentation review. The form completion required the engineer to thoroughly document the materials supplied by the IOU, assess the quality of the project documentation and savings estimates, and review baseline, EUL, costs, and other relevant information prior to the development of the site-specific M&V plan and site visit.

4.3.3 Site-Specific Measurement and Verification Plan Preparation

After completion of the lower rigor form, the evaluation engineer developed a Site-Specific Measurement and Verification Plan (SSMVP) for each gross impact sample point. This plan outlined the general ex-post impact approach, the IPMVP option, specification of the savings analysis method, and data collection requirements supporting the ex-post analysis, including the extent of monitoring efforts. The SSMVP also provided an analysis of the ex-ante methods and savings calculations to facilitate a discrepancy analysis focused on explaining the primary reasons for differences between the ex-post and ex-ante impact estimates.

4.3.4 On-Site Site Visits and Data Collection

On-site visits and data collection efforts were designed to verify equipment installation and operation and develop the best possible ex-post savings estimates for the budget and resources available. The process was designed to comply with the guidance documents referenced in this section.

Many of the SSMVPs were designed in compliance with IPMVP[8] Option A or B protocols that allow savings calculations based on short-term measurements of isolated systems. IPMVP Option A – partially measured retrofit isolation – incorporates verification and use of field measurement for some of the modeled parameters, while other parameters may be stipulated.

8 http://www.evo-world.org/index.php?lang=en

IPMVP Option B – retrofit isolation – is more rigorous and determines savings using field measurement for the systems affected by the measure in question.

Some projects were more appropriately handled with the IPMVP Option C (whole facility billing analysis using metering data) or Option D (calibrated building simulation). Option D was most often utilized in whole building and new construction analyses involving comprehensive HVAC system design upgrades and combined building system and/or envelope improvements.

In compliance with established guidelines and best practices, field measurements were collected whenever possible. This was an important step in the process because operating conditions found in the gross impact process may differ from those used in the ex-ante calculations supporting the IOU savings claims, and this difference in operating conditions often explains the discrepancy between ex-ante and ex-post savings. Often, data collected during the evaluation efforts clarified and revised hours of use, production rates, equipment loads, and other variables used to estimate savings.

Obtaining pre-retrofit and post-retrofit data for a full year was a goal for many of the projects. For several of the larger, more sophisticated participants, SCADA data served as the source for variables such as temperatures, flows, pressures, etc., as well as process throughputs. Challenges occurred when the required data were not available. Sometimes, participants did not monitor the specific parameter, or EMS trending capabilities were not enabled by the participant for that data point. In other cases unreliable data were obtained, including instances in which EMS or SCADA system sensor placement or calibration issues were uncovered, or where evaluation or program monitoring equipment malfunctioned. Whole site metering data was usually available; however, these data often did not provide the granularity required to analyze individual projects on large sites or to analyze demand impacts. Often billing analyses and/or use of calibrated models using whole site metering was not considered to be a reliable enough M&V solution due to such factors as the size of the expected impact relative to usage, variability in facility operations and/or production levels, and master metered campuses with measures only affecting a portion of the building population/systems. The above mentioned challenges were overcome by conducting field measurements using temporary data acquisition equipment. Typically, measurement equipment was deployed on site for between two and eight weeks, depending on the variability in equipment operation, evaluation project schedule and budget constraints, and customer willingness to participate in metering efforts.

During the site visit, the evaluation team engineer met with a facility representative knowledgeable about the equipment and operations and asked a series of questions regarding operating schedules, location of equipment, and equipment operating practices. During the on-site survey, data identified in the SSMVP were collected, including: monitoring records (such as instantaneous spot watt measurements, measured fluid and gas temperatures, data from equipment logs, energy management system (EMS) downloads, and SCADA system data), equipment nameplate data, and system operation sequences and operating schedules.

In many cases, the engineer would collect additional data using spot readings and/or temporary metering for power and/or runtime measurements on specific pieces of on-site equipment. In the case of temporary metering, the deployment of data collection equipment involved a second follow-up site visit or, in selected cases, the collection and return of field deployed monitoring equipment by a site contact.

During the site visit, the engineer also collected information through participant documentation and interviews with site personnel (often engineering or operations staff). This information related to the type, efficiency, and condition of pre-existing (replaced) equipment, effective useful life of installed equipment, remaining useful life of replaced equipment, project costs, and drivers for project implementation (other than energy savings). This information was used to establish project baseline efficiency and remaining useful life for gross impact calculations and was used to support net impact estimation (conducted by the NTG team).

Discrepancies were noted between the data requirements outlined in the SSMVP and actual field data collection performed, and the SSMVPs were revised as needed to reflect actual site conditions and M&V accomplishments.

4.3.5 Baseline Determination Procedures

Baseline determination is critically important in determining the program-induced energy savings and is described more fully in Appendix C. Code standard baselines (California Title 20 appliance standards, Title 24 building efficiency standards, or applicable Federal standards) were used to establish gross energy savings for normal replacement projects. For projects not addressed by code, industry standard practice (ISP) was used to establish the baseline. Industry Standard Practice, or ISP, is a term used to describe a technology or measure that is the typical equipment or commonly-used practice.[9] As necessary, based on the type of event, when projects did not identify a clear industry standard practice a survey was conducted among industry design professionals to investigate standard practice.

At times, determination of industry and company standard practice for specific measures was complex and required considerable effort to differentiate ISP from company standard practice.

---

[9] ISP Guide, Version 1.2A. http://www.cpuc.ca.gov/PUC/energy/Energy+Efficiency/Ex+Ante+Review+Custom+Process+Guidance+Documents.htm

- For example, assessment of industrial boiler combustion efficiency was particularly challenging because of the huge variation in size among facilities. In addition, boilers can be package units or field-erected units, and a large number of options exist for efficiency improvements related to combustion efficiency through the use of economizers, variable speed drives, oxygen flue gas trim systems, and so on. California Title 20 and Title 24 codes do not address the efficiency of large industrial boilers and neither do Federal regulations. Consequently, choosing appropriate baselines for industrial boiler projects required careful consideration. These types of baseline determinations were handled on a case-by-case basis.

- Another example involves a specific baseline issue found in wastewater treatment plants. Several large projects included the use of efficient single stage centrifugal blowers in aeration processes and variable speed drives in other applications in new design applications. Discussion with design professionals revealed that, for applications above approximately 900 horsepower, single stage centrifugal blowers are standard practice. In the case of variable speed drives, professionals surveyed stated that in new plant design these drives were incorporated in applications above approximately 3 horsepower, with the exception of wash water pumps.

The specification of project baselines is a particularly complex issue in cases where an Industry Standard Practice (ISP) baseline is being considered for evaluated projects. Under some circumstances, there is the potential to double-count the counterfactual effect, that is, what would have happened in the absence of the program, on the gross and net sides. If ISP baselines are set at too high an efficiency level, than double counting with NTG adjustments can occur. Conversely, setting all gross baselines at code or minimums that are well below non-participant market average baselines, absent the program, can result in overestimating savings. Actions that our evaluation team took to avoid the former were to set ISP conservatively, exclude the partial free ridership adjustment due to intermediate efficiency from the NTG algorithm,[10] and to investigate any cases in which ISP or dual baseline approaches resulted in low gross savings and were also associated with low NTGs. All such cases received careful reviews by the net impact team to assess occurrences of potential double counting. Where double counting was assessed to have occurred, such projects were eliminated from the net sample. A more complete description of this process is provided in Section 4.4.6 below.

[10] Self report-based, net-to-gross algorithms of the type used in this study often include an adjustment for "partial free ridership" associated with the participant's likelihood of adopting an intermediate efficiency level below that adopted through the program but above code or the market minimum. This adjustment was excluded in our NTG algorithm to avoid double counting due to the fact that baseline efficiency levels were set on the gross impact side of the analysis.

4.3.6 Prepare Ex-post Savings Calculations

After the field data collection process, the evaluation team engineer developed gas and electric savings based on data collected in the previous steps. Additional survey data were also used when industry standard practice investigations were undertaken to assist in baseline determination for gross savings calculations.

Energy savings calculations were performed using various methodologies including building energy simulation modeling, bin method calculations, engineering algorithms, analysis of pre- and post-installation billing and interval data, and other specialized technology or process-specific algorithms and models, as appropriate to the project. Peak demand savings were estimated for all projects with electric savings in the sample. Peak demand savings were calculated using a combination of billing and interval data, building simulation models with hourly outputs, and engineering calculations based on estimates of operating profiles and coincident peak diversity factors.

In certain cases, the analysis required additional data or confirmation from the site contact, from the IOUs, or from other implementing parties. These data were requested at the time of the site visit or, in some cases, after the site visit was complete. Specific follow up data requests were made to facilitate this process.

4.3.7 Final Site Report Preparation and Quality Control Review

The data collection and analysis activities described above were documented in a Final Site Report (FSR). The FSR detailed measure installation verification, as-found equipment operation, and final savings estimation.

In the FSR, the evaluators compared the ex-post savings as determined by the evaluating engineer against ex-ante savings supplied in the IOU documentation, and a GRR was calculated. The FSR also compared the baseline used in the IOU analysis with the most appropriate baseline as determined by the evaluating engineer. The FSR also showed the project type (for example, early replacement, normal replacement, replace on burnout, new construction, capacity expansion, system optimization, or add-on measure) determined by the evaluators and contrasted this with project type used for the IOU claim. In the FSR, the evaluators documented and detailed the reasons for discrepancies between claimed and evaluated savings. The FSR also listed and discussed effective useful life and remaining useful life, and noted any discrepancies in the tracking data and project documentation.

To ensure accurate analysis and reporting, the evaluating engineer and a reviewing engineer performed a quality review check of the FSRs and savings calculations, including the supporting field-collected data (equipment make, model number, capacity data, operation, photographs, nameplate data, screen shots, measurement data, etc.).

4.4 Net Impact Evaluation Methodology

The net impact methodology involved a three-step process:

- First, a net-of-free-ridership ratio was estimated for each site evaluated through analysis of surveys and/or interviews.
- Next, for those sites reporting spillover that was significantly program-influenced,[11] associated spillover savings were quantified and reported separately from the free ridership-based NTG results.[12]
- Third, a net-of-free ridership estimate was developed for the population by extrapolating from the sample to the entire population sample frame.

4.4.1 Overview of Approach to Estimating Free Ridership

The methodology described in this section was developed to address the unique needs of large nonresidential customer projects developed through energy efficiency programs offered by the four California IOUs and third-party implementers. As discussed further in the research plan for this evaluation and the evaluation guidance documents previously referenced, the free ridership method used for this study relies exclusively on the self-report approach (SRA) to estimate project and domain-level net-to-gross ratios (NTGRs), since other available methods and research designs are not feasible for the types of large nonresidential custom programs that were the subject of this evaluation.[13] The SRA in this evaluation was implemented in accordance with the relevant EM&V guidelines (see Appendix C) including the California Energy Efficiency Evaluation Protocols (April 2006).

---

[11] As indicated by a program importance score of 8, 9 or 10 in spillover decision making. NTG scoring is described below in section 4.4.4.

[12] Note that the quantification of spillover is still in process. Only a small number of projects, seven in total, met the requirements for spillover estimation.

[13] In the industrial sector, there are numerous obstacles to use of non-SRA methods, such as billing analysis and discrete choice methods, three of which are noted here. First, there is a signal to noise ratio challenge (low statistical power) in a participant/nonparticipant billing analysis, i.e., the expected difference in monthly energy use between participants and nonparticipants is difficult to detect reliably compared to other sources of variation in kWh across individual industrial sites due to the heterogeneity of the efficiency projects and large potential swings in baseline consumption due to non-efficiency factors. Second, some large industrial customers targeted by the program may have been influenced by participation in energy efficiency programs in prior years, making it very difficult to find true nonparticipants. Finally, even if the first two problems were absent, the large industrial customers targeted by the program are each relatively unique, making it unlikely that one could find a group of nonparticipants that could be matched with participants on critical variables to control for non-program factors.

The SRA methodology developed for this study provides a standard framework, including decision rules, for integrating findings from both quantitative and qualitative information in the calculation of the NTGR in a systematic and consistent manner. This approach was designed to fully comply with the CPUC Protocols and NTGR Guidelines referenced previously.

The method uses a 0 to 10 scoring system for key questions used to estimate the NTGR rather than using fixed categories that were assigned weights. Respondents were asked to jointly consider and rate the importance of the many likely events or factors that may have influenced their energy efficiency decision making for the project in question, rather than focusing narrowly on only their rating of the program's importance. This question structure more accurately reflects the complex nature of the real-world decision making and helped to ensure that all non-program influences were taken into account in assessing the unique contribution of the program to the energy efficiency project's implementation.

There are three levels of free-ridership analysis. The most detailed level of analysis, the *Standard – Very Large Project* NTG analysis, was applied to the largest and most complex projects with the greatest expected levels of gross savings[14] and incentives of greater than $200,000 (representing 10 to 20 percent of the total ex-ante domain impacts). The *Standard NTG* analysis, involving a somewhat less detailed level of analysis, was applied to projects with moderately high levels of gross savings, generally those with incentives of between $50,000 and $200,000. The least detailed analysis, the *Basic NTG* analysis, was applied to all remaining projects, generally those with incentives less than $50,000. The subsection below and Appendix D provide additional information on these free ridership levels of analysis.

4.4.2 NTG Rigor Levels

Table 4-1 below shows the data sources that were used in each of the three levels of rigor in the free-ridership analysis. Although more than one level of rigor may share the same source, the amount of information that was utilized in the analysis varied. For example, all three levels of rigor obtained core question data from the survey. The Large Nonresidential NTG Survey Instrument (also known as Decision Maker Survey) can be found in Appendix D.

[14] Note that we do not refer to an Enhanced level of analysis, since this is defined by the Protocols to involve the application of two separate analysis approaches, such as billing analysis or discrete choice modeling.

Table 4-1: Information Sources for Three Levels of NTGR Analysis

Level of Analysis	Program File	Decision Maker Survey Core Question	Vendor Surveys	Decision Maker Survey Supplemental Questions	Utility & Program Staff Interviews	Other Research Findings
Basic NTGR	√	√	√[1]		√[2]	
Standard NTGR	√	√	√[1]	√	√	
Standard NTGR - Very Large Projects	√	√	√[3]	√	√	√

Footnotes below reference question numbers provided in Appendix D.

1 Only performed for sites that indicate a vendor influence score (N3d) greater than maximum of the other program element scores (N3b, N3c, N3g, N3h, N3l).

2 Only performed for sites that have a utility account representative.

3 Only performed if significant vendor influence was reported or if secondary research indicated the installed measure may be becoming standard practice.

4.4.3 NTG Survey Process

The project data collection process for the net-to-gross effort began with a data request to each IOU for additional information to supplement the data provided through the tracking database. The data request sought name and contact information for the primary decision maker and the account executive, a listing of program services received, and a more comprehensive project and/or measure description. All of this information was needed to ensure that the NTG team had a comprehensive understanding of the project and related program interventions and was able to efficiently contact the correct decision maker.

All of the gross impact and LRA sample points were targeted for inclusion in the NTG sample, and a large number of the associated net-to-gross interviews were completed with the participants' energy efficiency project decision makers. Additional NTG points were also included to improve precision levels and to allow for NTG estimation at the program group level. NTG interviews were conducted by both professional consulting staff and by staff in Itron's computer-aided telephone interviewing (CATI) center, depending on the rigor level of the sample point. For the NTG sample points that overlapped with the gross impact sample, the interviewing team worked with the engineering team to prepare for the NTG survey. As noted above, following the completion of these surveys, information from both the gross and net activities was considered when determining the final net savings estimates. Care was taken to ensure that the results were internally consistent and did not include any double counting of effects between the two samples. Detailed site-specific NTG results can be found in Appendix D-3.

4.4.4 NTG Questions and Scoring Algorithm

The NTGR was calculated as an average of three scores. Each of these scores represents the highest response or the average of several responses given to one or more questions about the decision to install a program measure.

1. S**core 1** that reflects the influence of the **most important** of various program and non-program elements in the customer's decision to select the specific program measure at this time. Program influence through vendor recommendations was also incorporated in this score.

2. S**core 2** that captures the perceived importance of the program (whether incentive, recommendation, training, or other program intervention) relative to non-program factors in the decision to implement the specific measure that was eventually adopted or installed. This score was determined by asking respondents to assign importance values to both the program and most important non-program influences so that the two values total 10. The program influence score was adjusted (divided by 2) if respondents said

they had already made their decision to install the specific program qualifying measure before they learned about the program.

3. S**core 3** that captures the likelihood of various actions the customer might have taken at the time or project decision making, and in the future, if the program had not been available (the counterfactual). This score also accounts for deferred free ridership by incorporating the likelihood that the customer would have installed program-qualifying measures at a later date if the program had not been available.

When there were multiple questions that fed into the scoring algorithm, as was the case for **Score 1**, the maximum value for program and non-program influences was always used. The rationale for using the maximum value was to capture the most important program element in the participant's decision making. Thus, each score was always based on the strongest influence indicated by the respondent. However, high scores that were inconsistent with other previous responses triggered consistency checks and led to follow-up questions to clarify and resolve the discrepancy. Note that Score 1 took the highest program score divided by the sum of the maximum of the program and non-program scores.

When there were missing data or 'don't knows' to critical elements of each score, one of two options was used. The most common approach, in cases where it was one of several other elements that were considered in the algorithm, was to simply exclude the missing element from consideration.

The resulting self-reported NTGR in most cases was simply the average of all three scores, divided by 10. The one exception to this was when the respondent indicated a 10 in 10 probability of installing the same equipment at the same time in the absence of the program, in which case the NTGR was based on the average of Scores 2 and 3 only.

4.4.5 Data Analysis and Integration

The calculation of the NTGR for Basic and Standard rigor projects (i.e., those with incentives of $200,000 and less) was generally mechanical and was based on the answers to the closed-ended questions. The calculation of the NTGR for Standard - Very Large rigor projects (with incentives over $200,000) relied on additional information from other sources and used a more detailed 'case study' level of effort in many cases. The SRA guidelines point out that a case study is one method of assessing both quantitative and qualitative data in estimating a NTGR. A case study is an organized presentation of available data about a particular customer project with respect to all relevant aspects of the decision to install the efficient equipment. In such cases where multiple interviews were conducted, eliciting both quantitative and qualitative data and a variety of program documentation, all of this information was integrated into an internally consistent and coherent story that supported a specific NTGR. This process, in which multiple data sources are used to develop the NTGR, is referred to as "triangulation" and was used for a small number of Standard-Very Large projects.

Sometimes, *all* the quantitative and qualitative data clearly pointed in the same direction while, in others, the *preponderance* of the data pointed in the same direction.[15] Other cases were more ambiguous. For all Standard-Very Large projects, in order to maximize reliability, it was essential that more than one person be involved in analyzing the data. Each person analyzed the data separately and then compared and discussed the results. Important insights can emerge from the different ways in which two analysts look at the same set of data. Ultimately, differences were resolved and a case made for a particular NTGR. Careful training of analysts in the systematic use of rules was carried out to insure inter-rater reliability.[16]

Once the individual analysts completed their review, they discussed their respective findings and presented their respective rationales for any recommended changes to the equation-derived NTGR. The outcome of this discussion was the final NTGR for a specific project.

4.4.6 Adjustment to Net Sample to Account for Overlapping ISP and Dual Baseline Sites

For that portion of the Net-to-Gross (NTG) sample that overlapped with the gross sample, several projects were eliminated to avoid possible double-counting between the gross and net results (the situation where project realization rates are reduced on both gross and net for the same reason). *Note that this was for the limited purpose of calculating an NTGR Adjustment Factor*. Prior to the draft report submittal, two projects had been removed from the net sample. These two projects had zero gross savings and very low NTGRs, and were presumed to be cases of double counting.

Following the completion of the draft report, a second review was completed to discern additional projects with potential double counting. This review consisted of projects that had non-zero ex-post gross savings results. In all cases, the evidence behind each project was carefully examined, including an in-depth review of each site report. The following process was then used to eliminate these overlapping projects:

- **ISP Baseline projects.** Sites were identified for removal that met the following criteria: (1) the ex-post baseline disposition was Industry Standard Practice [ISP], (2) the primary reason for the discrepancy was due to the assumption of an ISP baseline as standard, (3) the Gross Realization Rate (GRR) was low (0.35 and below) and the NTGR was low

15 These cases resulted in NTG scores that tended to be close to 0 or 1.

16 Inter-rater reliability is the extent to which two or more individuals (coders or raters) agree. Inter-rater reliability addresses the consistency of the implementation of a rating system.

(0.30 and below). Most of these sites also claimed they would have installed the same measure in the program's absence suggesting a high likelihood of double counting. One site was a duplicate of another claimed project; it was also dropped as an ineligible measure. A total of five projects were dropped for these reasons.

- **Dual baseline projects.** Sites were identified for removal that met the following criteria: (1) the Remaining Useful Life (RUL) was short, (2) the GRR in the ex-post RUL-EUL period was low or zero, and (3) the NTGR was low (0.30 and below). A total of four projects were dropped for these reasons. There were also a few projects that met the RUL and post-RUL GRR thresholds but had high NTGRs (0.60 and above); those remain in-sample. This is to give the benefit of the doubt to the program in such cases.
- These nine overlapping projects were then removed from the NTGR sample frame. A total of 266 projects overlapped between the gross and net samples prior to this removal, and 257 projects overlapped after this removal.
- Following this, the NTGRs were re-run for the overlapping project population only. First, NTGRs were calculated for the overlapping points with the nine projects included (to establish base values). Next, NTGRs were calculated for the overlapping points with the nine projects removed. The resulting NTGRs were then compared, and the percentage increase between the base and non-overlap cases was computed.
- A multiplier of 1 plus this percentage increase was then developed for each IOU-fuel sampling domain. This multiplier was applied to the NTGR values from the draft report to obtain revised NTGRs for the final reporting of results. Note that only the IOU-fuel domain NTGRs were adjusted; results were not re-run at the stratum or program level because in some cases the sample sizes were not sufficient. *In all cases, the improvement in IOU-fuel domain level NTGRs was very slight, on the order of 1 to 2% for all IOU-fuel domains except PG&E Gas (which had zero projects removed and therefore didn't change).*

4.5 Lower Rigor Assessment Evaluation

As introduced previously, additional project level evaluation efforts were conducted using a lower rigor assessment approach for an expanded sample of all gross impact points and supplemental 'LRA only' points. Supplemental points were strategically allocated to custom programs to provide a sufficient LRA sample size to support an impact-oriented assessment of specific programs and program groups. The selection of custom programs was a collaborative process between the CPUC staff, the Custom Impact (WO33) evaluation team, and the IOUs.[17]

[17] PG&E and SCE constructed a list of programs they identified as programs of interest.

The LRA approach was intended to provide cost effective, qualitative, impact-oriented findings and feedback for a broad range of programs. This approach involved a desk review of project application documentation and completion of a standard LRA form. The LRA involved a review of project application paperwork received from the IOU, an assessment of the quality of the documentation provided, a check for adherence to rules and guidelines, and an assessment of savings estimation techniques.[18] The goal of this effort was to provide a broad qualitative assessment of the successes and shortcomings of the program implementation processes. Lower rigor points expanded the reach of the evaluation to programs that would otherwise not receive much attention based on the gross impact sample allocation alone, due to budget constraints. Lower rigor assessment results did not factor into to the determination of M&V-based, ex-post custom impact energy savings; they provided more general feedback regarding conformance with sound impact-related and project application practices for projects and programs.

The lower rigor assessments were completed using a template form that guided the evaluating engineer through critical application quality issues. These issues were selected because they were considered to be important to a thorough and complete evaluation, as well as to reflect problems that were flagged through the evaluation process of the previous 2006-2008 custom programs. The LRA assessments addressed the following questions:

- Was the IOU application complete and accurate?
- Was the IOU tracking data complete and accurate?
- Did the project utilize pre-installation M&V?
- Was the appropriate baseline specified?
- Was a valid RUL / EUL approach used (for early replacement claims)?
- Was an appropriate impact calculation method used?
- Were all relevant inputs considered, i.e., production levels?
- Were there adequate values for all inputs?
- Was an appropriate HVAC interactive effects calculation method used?
- Was an appropriate non-HVAC interactive effects calculation method used?
- Did the project utilize post-installation M&V?
- Were the measures eligible?
- Did the measures exceed code or industry standard practice?
- Were multiple IOU fuel impacts properly accounted for?

[18] See Appendix E for more information on the details of the Lower Rigor Assessment Form and an explanation of issues assessed.

- If applicable, was fuel switching supported with the three prong test?[19]
- Were the non-IOU fuel and ancillary impacts of the project properly accounted for (cogeneration/waste heat recovery/ refinery gas, etc.)?
- Did the customer installation meet all program rules?

These critical issues were assessed as follows:

- Assessable (Yes/No/NA):
  - Yes, documentation was adequate to allow for an assessment
  - No, not enough information available in the project files to make an assessment
  - Not applicable (this issue does not apply to the particular project being reviewed)
- Quality (good, neutral, poor):
  - Good: the treatment of this issue clearly meets protocol and quality guidelines
  - Neutral: the treatment of this issue isn't clearly flawed and isn't clearly well within quality standards
  - Poor: the treatment of this issue does not meet protocol and/or quality guidelines for project applications

Following the completion of the LRA forms, the results were tabulated and scored across categories. Results were combined on a program or program group level; trends and differences were highlighted. LRA results were analyzed by the following categories or groupings:

- IOU
- IOU fuel domains
- IOU-specific program groups (sampling domains)
- Statewide program groups (sampling domains)
- Individual programs

Results of the LRA are provided in Chapter 7; additional results, details regarding the assessment template, and scoring criteria are presented in Appendix E.

[19] Fuel substitution projects must pass the three-prong test to ensure that source BTU consumption does not increase and that the project does not cause adverse environmental impacts, while passing applicable benefit-cost tests, as described in the Energy Efficiency Policy Manual (http://www.cpuc.ca.gov/NR/rdonlyres/7E3A4773-6D35-4D21-A7A2-9895C1E04A01/0/EEPolicyManualV5forPDF.pdf).

For the gross impact points, the LRA form was completed prior to beginning site work. The form thus captured the initial assessment based on IOU project documentation alone, rather than subsequent M&V. For the BD period, the LRA was only based on the pre-M&V information. For the AD period, the LRA was also first conducted prior to the M&V activities; however, the LRA for the AD period was subsequently adjusted based on the M&V results. This post M&V LRA adjustment was an activity added to the project scope after completion of the BD sites in order to compare the LRA with and without M&V results. The LRA form required additional notations when the assessment had changed as a result of M&V activities for the AD gross impact sites. The conclusions based on pure application (desk) review were compared with the conclusions obtained with the benefit of telephone contact, site visits, additional data collection, and more thorough analysis. Examination of the reasons for the change in assessment and improved project understanding can be used to assess the value of these LRA activities. The point in the gross impact estimation process where the discrepancy was noted was recorded using the list provided below:

- Desk review QA/QC
- IOU data request
- Customer recruitment contact
- Customer follow-up contact
- IOU account rep phone call
- First on-site visit
- Second on-site visit
- NTG interview
- Savings analysis calculations
- Final site report drafting
- Final site report QA review

5

Gross Impact Results

This chapter presents quantitative and qualitative gross impact results for the 2010-2012 WO033 custom impact evaluation. Gross impact realization rates (GRRs) are presented in this chapter using a variety of segments and combinations of those segments, including results by project, IOU fuel domain, size stratification, program group and measure. Results are also presented for energy metrics – electric energy (kWh), electric demand (kW) and gas energy (therms).

Sampling domains are defined in the WO033 research plan[1] and in Chapter 3. Briefly, these domains include five combinations of IOU and fuel: PG&E electric, PG&E gas, SCE electric, SDG&E electric, and SCG/SDG&E gas. Unless noted otherwise, realization rates represent the full lifecycle of the projects examined, that is, the lifecycle ex-post evaluation-based estimate of impacts divided by each the IOUs' lifecycle ex-ante estimate of impacts.

5.1 Project-Specific Gross Impact Summary

Weighted gross impact results are presented in this section by IOU fuel domain and stratum. Gross impact evaluation results are supported by 429 M&V sample points. A sample point can include more than one tracking system record and/or consist of an aggregation of IOU applications. In addition, some gross impact points include both ex-ante electric and gas savings in the PG&E and SDG&E domains. The 429 impact sample points account for 495 electric and gas fuel-differentiated results. This is because some sample points include claims for both electric and gas savings. These 495 results are referred to in this section as "projects."

Overall, the custom impact evaluation includes results for 316 electric projects and 179 gas projects. Figure 5-1 and Figure 5-2 graphically display ex-post versus ex-ante savings estimates for the statewide sample. These figures present the ex-ante (tracking system) claimed savings and the ex-post evaluated savings for the M&V sample points for kWh and therms, respectively. The charts also include a unity line, which divides the results into those in which the project-specific realization rates were above one (sites above the line) and below one (sites below the line). Electric projects with ex-ante or ex-post savings greater than 7,250,000 kWh and gas projects with savings greater than 3,500,000 therms are excluded from the scatter plots below for

[1] http://www.energydataweb.com/cpucFiles/pdaDocs/814/WO33%20Research%20Plan%20Final%2012%2029.pdf

readability. The plots also do not include two projects with extreme realization rates, again to maintain readability for the bulk of the results. The figures show the majority of projects and include 99 percent of the 316 electric projects (312 projects) and 99 percent of the 179 gas projects (177 projects).

Figure 5-1: Lifecycle Ex-Ante and Ex-Post Electric Savings (kWh) for Sampled Projects (Extreme Ratios and Extremely Large Projects Removed)

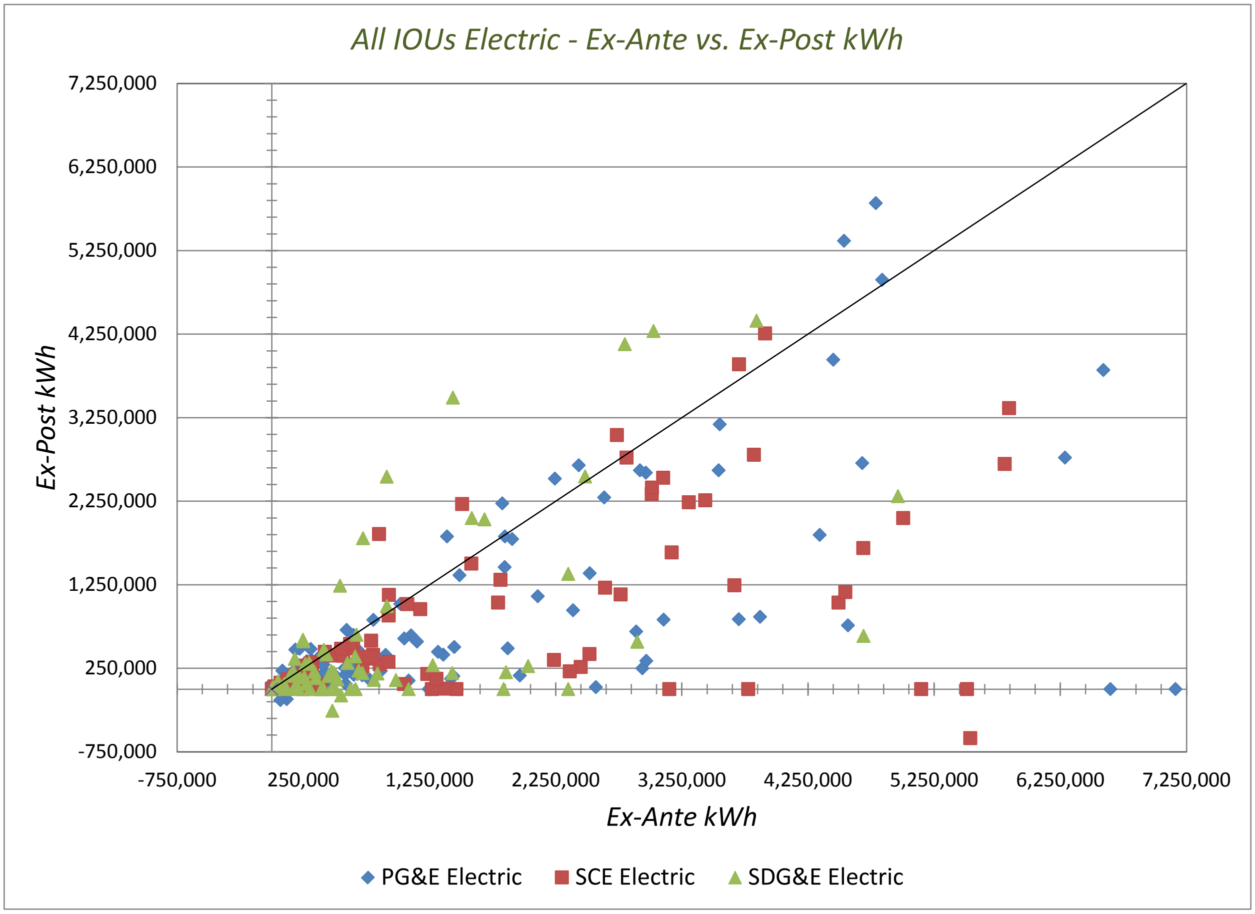
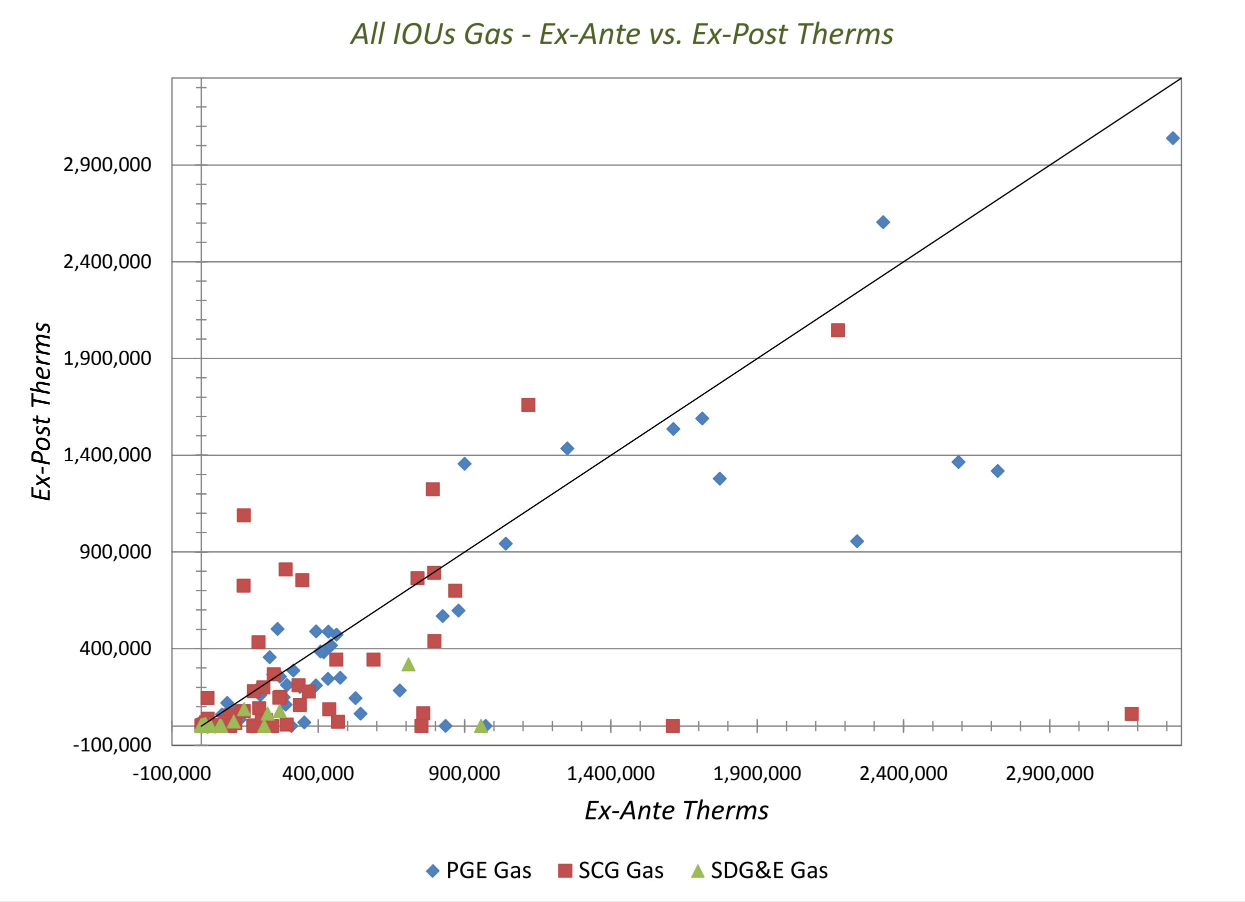


Figure 5-2: Lifecycle Ex-Ante and Ex-Post Gas Savings (Therms) for Sampled Projects (Extreme Ratios and Extremely Large Projects Removed)



Very few of the projects yielded GRRs that exceed one and lie above the line in the chart. As shown in the tables that follow, despite higher than planned error ratios (ERs), weighted average GRRs by IOU fuel domain tend to be statistically significantly less than one and greater than zero.

Mean gross realization rates for all IOUs for the appropriate energy metrics (kW, kWh, or therms) are presented graphically in Figure 5-3 and Figure 5-4. Figure 5-4 shows the results with observations associated with extreme GRRs removed. GRR results with extreme points excluded are provided for assessment and comparison purposes only. Evaluation savings of record includes all M&V points, including extreme GRRs, as the projects with extreme GRRs are not associated with evaluation measurement error.

Figure 5-3: Gross Realization Rate for All IOU Fuel Domains (All Completed Points)

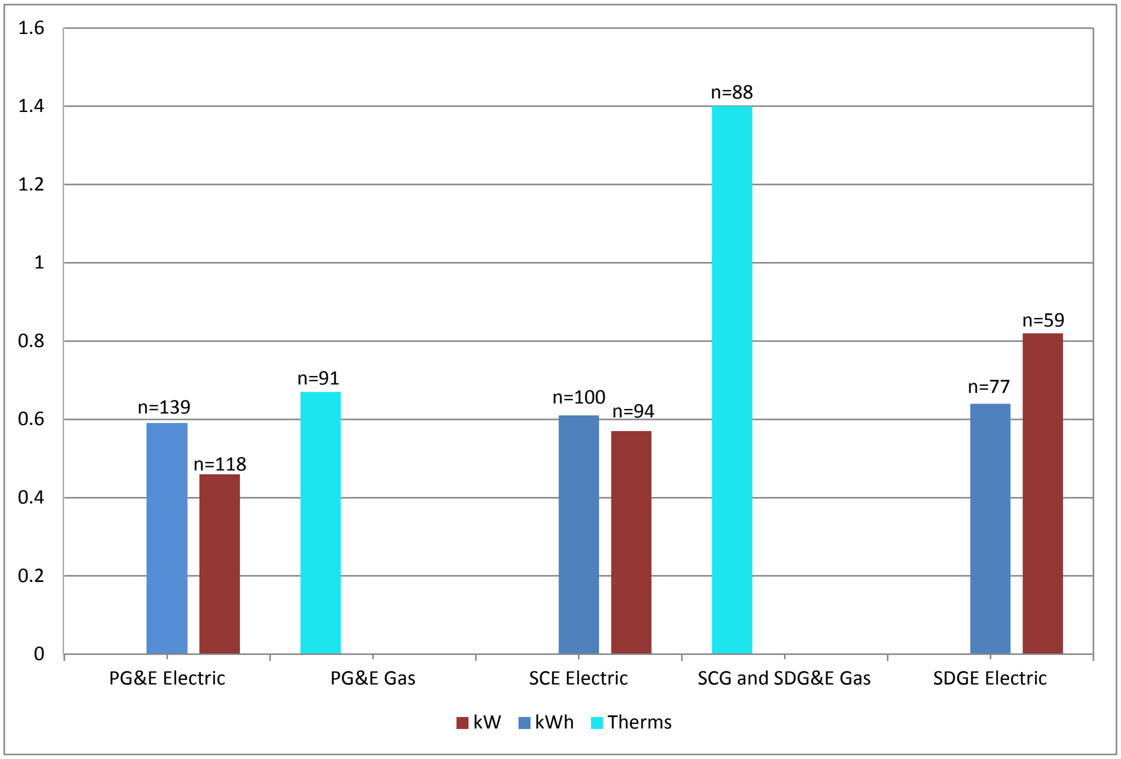


Figure 5-4: Gross Realization Rate for All IOU Fuel Domains (Excludes Projects with Most Extreme GRRs)

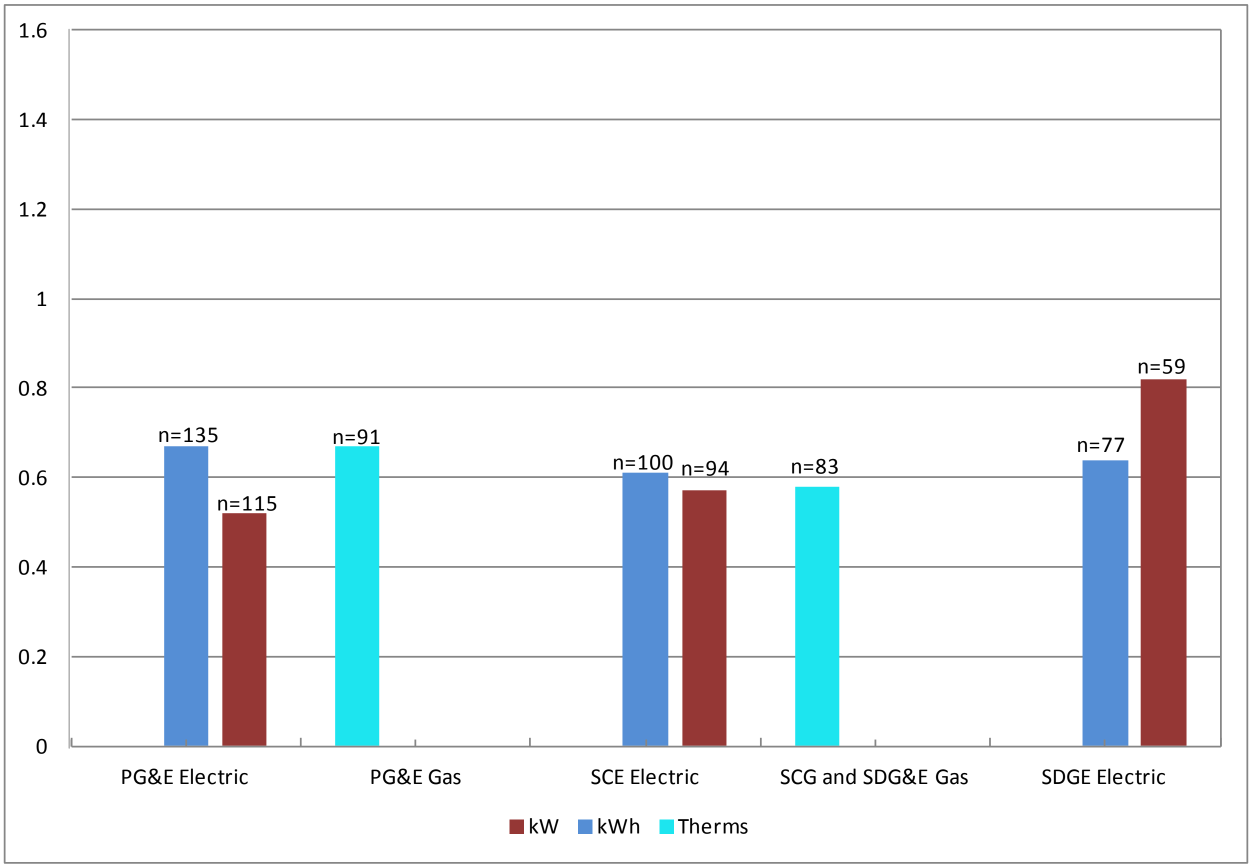


Table 5-1 presents project lifecycle (LC) GRRs for each of the five IOU fuel domains. The mean weighted realization rate is shown for kW, kWh and therms as a separate row for each domain, and indicates the frequency of realization rates that are higher than 150 percent, lower than zero percent (signifying an energy penalty), and equal to zero percent (signifying no energy savings). The population sample frame and the total number of completed gross impact points is also shown for each energy metric, along with the resulting ER. Lastly, first year (FY) GRRs are presented for comparison purposes. Table 5-1 includes all sample points; extreme GRRs are not removed.

With all GRRs included, the error ratio for most domains is considerably higher than the 0.80 error ratio expected based on past evaluations (primarily, the *2006-2008 Evaluation Report for PG&E Fabrication, Process and Manufacturing Contract Group*).[2] The mean realization rates by IOU, fuel and energy metric are less than 0.70 for all but two energy metrics. One domain, SCG/SDG&E gas, has a GRR that exceeds 1.00, while SDG&E kW has a GRR of 0.82. The

[2] http://www.calmac.org/publications/PG&E_Fab_06-08_Eval_Final_Report.pdf

results for the SCG/SDG&E gas domain are impacted by the calculations from three projects with extreme GRRs of 40, 46 and 53. All were sampled as SDG&E electric projects but had relatively small ex-ante gas impact claims that were estimated to be significantly larger in the ex-post evaluation.

Table 5-1: Weighted Project Lifecycle Gross Realization Rates by Sample Domain and Energy Metric (kWh, kW, and Therms) - All Completed Projects

Energy Metric	Sample Count	Population N	ER	LC GRR Mean	90% Confidence Interval	RR > 1.50	RR = 0	RR < 0	FY GRR Mean
PG&E Electric									
kWh*	139	6,994	1.26	0.59	0.49 / 0.70	9	10	4	0.61
kW	118	6,248	1.67	0.46	0.35 / 0.58	9	8	9	0.53
PG&E Gas									
Therms*	91	1,270	0.56	0.67	0.61 / 0.74	10	9	7	0.70
SCE Electric									
kWh*	100	3,052	1.03	0.61	0.51 / 0.71	3	14	1	0.60
kW	94	2,748	1.08	0.57	0.47 / 0.67	7	16	2	0.61
SDGE Electric									
kWh*	77	1,469	0.99	0.64	0.52 / 0.75	10	16	2	0.46
kW	59	790	2.10	0.82	0.46 / 1.17	14	15	2	0.84
SCG and SDG&E Gas									
Therms*	88	1,077	4.99	1.40	0.23 / 2.57	5	16	2	1.18

* Primary sample was designed and selected at this level. Note that the kWh and kW sample and population counts are not equal, as not every electric project included a kW saving claim

Table 5-2 presents weighted project LC GRRs for each IOU fuel domain, excluding projects with the most extreme high or low realization rates. Three of these projects with a GRR exceeding 30 are in the SDG&E/SCG gas domain, as noted above (H026, H032 and H401). Six other projects, four PG&E electric projects (E059, E091, E237 and E426) and two other SDG&E/SCG gas projects (G418 and H214) also have an extreme GRR, with a very low GRR (below -0.74) for the primary fuel evaluated. Sensitivity analysis was performed by removing the relatively few projects meeting these criteria, in order to assess the effect of projects with extreme realization rates on the average GRR. Results with extreme points removed better represent the majority of projects. A total of 12 domain and energy metric combinations (4 PG&E electric kWh, 3 PG&E electric kW, and 5 SDGE/SCG gas projects) were classified as extreme points. No ‘extreme GRR’ projects are present in the SCE and SDG&E electric domains. All twelve domain and energy metric combinations classified as extreme, which arise from nine sample points, are described in Appendix C along with the reasons for the discrepancies in GRR. Again, extreme GRR points are only excluded for comparison purposes. Evaluation savings of record includes all M&V points, including extreme GRRs.

Table 5-2: Project Lifecycle Realization Rates by Sample Domain and Energy Metric (kWh, kW, and Therms) - Excludes Projects with Most Extreme GRRs**

Energy Metric	Sample Count	Population N	ER	LC GRR Mean	90% Confidence Interval	RR > 1.50	RR = 0	RR < 0	FY GRR Mean
PG&E Electric									
kWh*	135	6,994	0.92	0.67	0.58 / 0.75	9	10	0	0.68
kW	115	6,248	1.29	0.52	0.41 / 0.62	10	8	6	0.58
PG&E Gas									
Therms*	91	1,270	0.56	0.67	0.61 / 0.74	10	9	7	0.70
SCE Electric									
kWh*	100	3,052	1.03	0.61	0.51 / 0.71	3	14	1	0.60
kW	94	2,748	1.08	0.57	0.47 / 0.67	7	16	2	0.61
SDGE Electric									
kWh*	77	1,469	0.99	0.64	0.52 / 0.75	10	16	2	0.46
kW	59	790	2.10	0.82	0.46 / 1.17	14	15	2	0.84
SCG and SDG&E Gas									
Therms*	83	1,077	0.80	0.58	0.50 / 0.66	8	16	2	0.64

* Primary sample was designed and selected at this level.

**Only shaded energy metric domains had projects with extreme GRRs removed.

After removing the projects with the most extreme GRRs, analysis of the GRRs across sample domains (IOU fuel) does not yield statistically significant differences (considering the corresponding confidence interval), which is not surprising given the confidence interval for the SCG/SDG&E gas domain shown in Table 5-1 (0.23 – 2.57).[3] The error ratio with outliers removed is not far from the evaluation team's research planning expectations, ranging between 0.56 and 1.03 for the primary fuel evaluated (electric kWh or gas therms). As would be expected, the error ratio for the SCG/SDG&E gas domain decreases dramatically from 4.99 to 0.80 with the removal of the extreme GRRs in that domain. The evaluation indicates that, even with the extreme GRR projects removed, there are a large number of points with gross realization rates greater than 150 percent, equal to 0 percent, or less than zero percent for the primary fuel evaluated (40, 65, and 12 points respectively, out of a total of 429 points). This spread, particularly on the lower side, contributes to the relatively high error ratios. The mean realization rates with outliers removed are similar to some of the 2006-2008 custom impact

[3] To obtain results with a higher level of statistical confidence and precision, ex-ante savings estimates would need to align more closely with ex-post estimates or sample sizes would need to increase from present values; however, increasing sample sizes is probably unlikely given the total evaluation resources typically available and the time and cost of intensive ex-post M&V for custom sites. There is considerable room for improvement in ex-ante impact estimation. Areas for improvement should emphasize better engineering review and incorporation of M&V data in the development of ex-ante impact estimates.

evaluations. A large number of projects with negative and/or zero GRRs in all domains served to lower the mean realization rate in each domain. The factors that brought about the lower realization rates – the reasons for discrepancy – are explored in the next section.

The impact sample design used project size to develop sampling strata. This is a common and very effective technique for increasing the statistical power of a given sample size for a population with extremely wide ranging impacts. Table 5-3 presents impact results by size strata for each sample domain (and includes all sample points). Each sample domain has five strata, based on the size of claimed ex-ante energy savings, with strata 1 projects claiming the largest savings and strata 5 projects claiming the smallest savings. Sample strata were chosen to meet overall sample design goals; they are not designed to be statistically significant in and of themselves. Please note that the sample sizes for each stratum are small and the stratum level results should be interpreted with caution; however, the results are illustrative of project size-related trends within the various sample domains. Our observations on Table 5-3 include the following:

- The stratum 1 populations for both the PG&E gas domain and the SCG/SDGE gas domain have only one project; in both cases the projects are large refinery waste heat recovery projects that were combined for weighting purposes with strata 2 projects due to the statistical requirement that a stratum must have at least two sample points. The individual project GRR for the largest PG&E gas project is 0.98 and the GRR for the largest project in the SCG/SDG&E gas domain is 0.33. This illustrates the importance to the overall domain-level GRR result that a single project can have, given that just one, two or three projects might be large enough to command an entire stratum, and each stratum has roughly equal weight in the result.
  - The fact that PG&E electric and gas domain projects achieved relatively high GRR results in the highest strata provides a level of insurance against a poor overall result. Obtaining accurate results for the largest and most influential projects has been a recommendation from previous custom and standard performance contract (SPC) program evaluations for a number of program cycles. *PG&E should continue, and the other IOUs should improve, their focus on estimation accuracy and quality control for the very largest projects.* For example, SCE electric results were dragged down by the 0.33 GRR in strata 1.
  - A countervailing result is illustrated by strata 5 in the PG&E electric domain, which has the low realization rate results in strata 3 and strata 5. *PG&E engineering/quality control/application procedures should better address small projects or, if there is IOU concern about whether that can be done cost effectively, minimum size requirements for custom projects should be considered*. This is likely further challenged by the large number of programs with custom offerings that PG&E operates, including third party programs and government programs. As

illustrated in Section 5.3, PG&E's core programs are outperforming these other offerings, indicating tighter controls may be in place for core offerings.

- The realization rate for kW in the SDG&E electric domain, strata 5, is 1.22; this is significantly different than the kWh realization rate of 0.49. A large difference between the tracking data and the IOU-supplied analysis report for two Savings by Design new construction projects is the cause for this divergence. The realization rate is calculated based on reported savings in the IOU tracking data which, for these projects, was much lower than the savings estimates shown in the project-specific documentation provided by the IOU. *Quality control should be implemented to ensure that tracking system impacts match IOU installation reports*.
- As discussed above, the impact of the three extreme points in the SCG/SDGE gas domain can be seen in stratum 5, resulting in a GRR of 4.96. Removing these extreme points results in a weighted GRR of 0.58 in strata 5 and a drop in the overall domain result from 1.40 to 0.58.

A summary of project-specific characteristics and results for each individual gross impact project is provided in Appendix C. These characteristics include the type of measure and site, the strata and fuel type, gross realization rates, the primary discrepancy factor, mapping of Itron site IDs to IOU Claim IDs or application numbers, ex-ante energy savings from the IOU tracking systems, the LRA score, and the net to gross ratio.

Table 5-3: Project Lifecycle Realization Rate by Strata Across Sample Domains (All Completed Sample Points)

IOU Fuel Domain	Strata	Project Count	RR Mean kW	RR Mean – kWh	RR Mean - Therms	RR > 150%	RR = 0%	RR < 0%
PG&E Electric	1	19	0.98	0.87	-	1	2	0
	2	24	0.53	0.60	-	1	0	0
	3	25	0.44	0.35	-	0	2	1
	4	28	0.39	0.68	-	3	2	0
	5	43	0.28	0.50	-	4	4	3
PG&E Gas	1	3	-	-	0.97	0	0	0
	2		-	-		0	0	0
	3	10	-	-	0.63	0	0	1
	4	32	-	-	0.64	3	3	0
	5	46	-	-	0.52	7	6	6
SCE Electric	1	15	0.33	0.33	-	0	3	1
	2	20	0.63	0.62	-	1	2	0
	3	22	1.05	0.90	-	2	3	0
	4	20	0.49	0.63	-	0	2	0
	5	23	0.46	0.58	-	0	4	0
SCG/SDG&E Gas	1	14	-	-	0.49	0	0	0
	2		-	-		0	3	0
	3	20	-	-	0.76	3	2	0
	4	22	-	-	0.63	1	5	0
	5	32	-	-	4.96	4	6	2
SDG&E Electric	1	13	1.08	0.65	-	1	2	0
	2	16	0.68	0.76	-	2	2	0
	3	16	0.48	0.39	-	1	4	2
	4	15	0.69	0.82	-	3	2	0
	5	17	1.22	0.49	-	3	6	0

5.2 Discrepancy Analysis

This section presents an analysis of the discrepancies that account for the difference in ex-ante and ex-post savings for the sampled projects. Note that this discrepancy analysis is based on discrepancies associated with first year impacts and gross realization rates.[4]

The evaluation identified seven important discrepancy factors[5] leading to adjustments to the ex-ante claims stemming from ex-post gross impact estimates in the M&V sample. The seven factors are: operating conditions, inappropriate baseline, calculation method, ineligible measure, equipment specification, measure count, and tracking database discrepancy. When examined for both the frequency and the degree of impact on the ex-ante savings claims, three factors are most influential (or dominant):

- Differences in operating conditions (for example, hours of operation, VSD speeds, return to original operation, operations are forecasted instead of using observed or measured conditions, production levels, etc.)[6]
- Inappropriate baselines or baseline conditions used for ex-ante savings estimation (for example, rejected early replacement claims, new equipment that do not exceed code-required efficiency levels, new equipment that do not exceed industry standard efficiency levels, rejected normal replacement claims, inaccurate baseline or pre-retrofit operating hours, etc.)
- Calculation methods used for ex-post savings estimation were revised from those used to estimate ex-ante savings (for example, simulation models that utilize Title 24 compliance schedules as opposed to actual building operation; switching from Energy Pro to eQuest building simulation software; different spreadsheet calculation approaches based on post-retrofit or post-construction data availability; expanded spreadsheet approaches to account for varying loads and interactive effects; shifts to or from the use of billing analyses and interval data, particularly for peak demand impacts; use of calculation inputs defining initial or ex-post operating conditions when not attributable to other discrepancy factors; etc.)

---

4 Thus, the effect of ex-post dual baseline adjustments on lifecycle GRRs is not reflected in this discrepancy analysis.

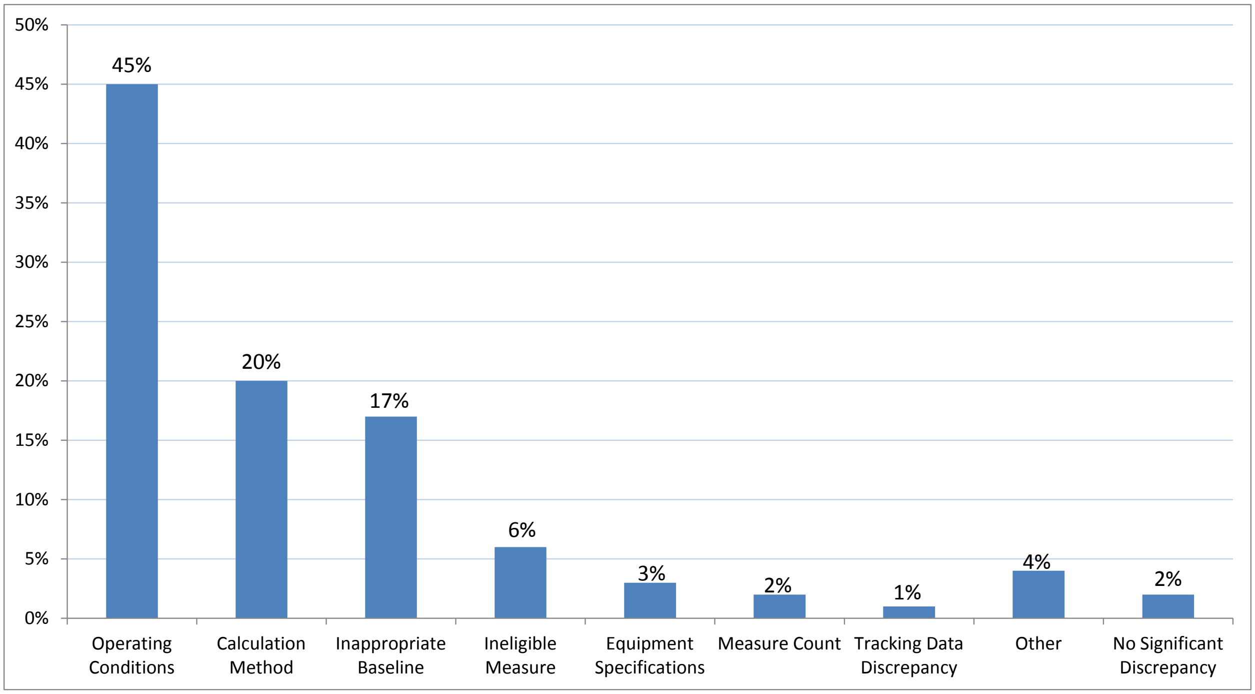
5 A separate ‘Other’ category includes less common factors and accounts for a relatively small number of projects and percentage of savings. These factors are inoperable measure, program rule compliance, measure not installed, and unquantified fuel impacts.

6 Operating conditions often change over time due to business conditions or other changes at a facility, and the IOUs can do little to control adjustments in operations after savings are claimed. In many instances, however, operating conditions had changed before the time of the IOU’s or implementer’s final inspection, but ex-ante savings were not always updated in such instances.

5.2.1 Frequency of Discrepancy Factors

Given multiple tracking records associated with some projects, 1,215 records associated with 495 projects are examined in this section. Each record was assigned a primary (and sometimes a secondary) factor that explains the observed discrepancy. The three most frequent discrepancies cited above account for 81 percent of all discrepancies found. These three discrepancies also account for a 35 percent shortfall from the kWh ex-ante claims and a 26 percent reduction in gas ex-ante claims. Figure 5-5 shows the frequency of these key factors as a percent of all records assessed.

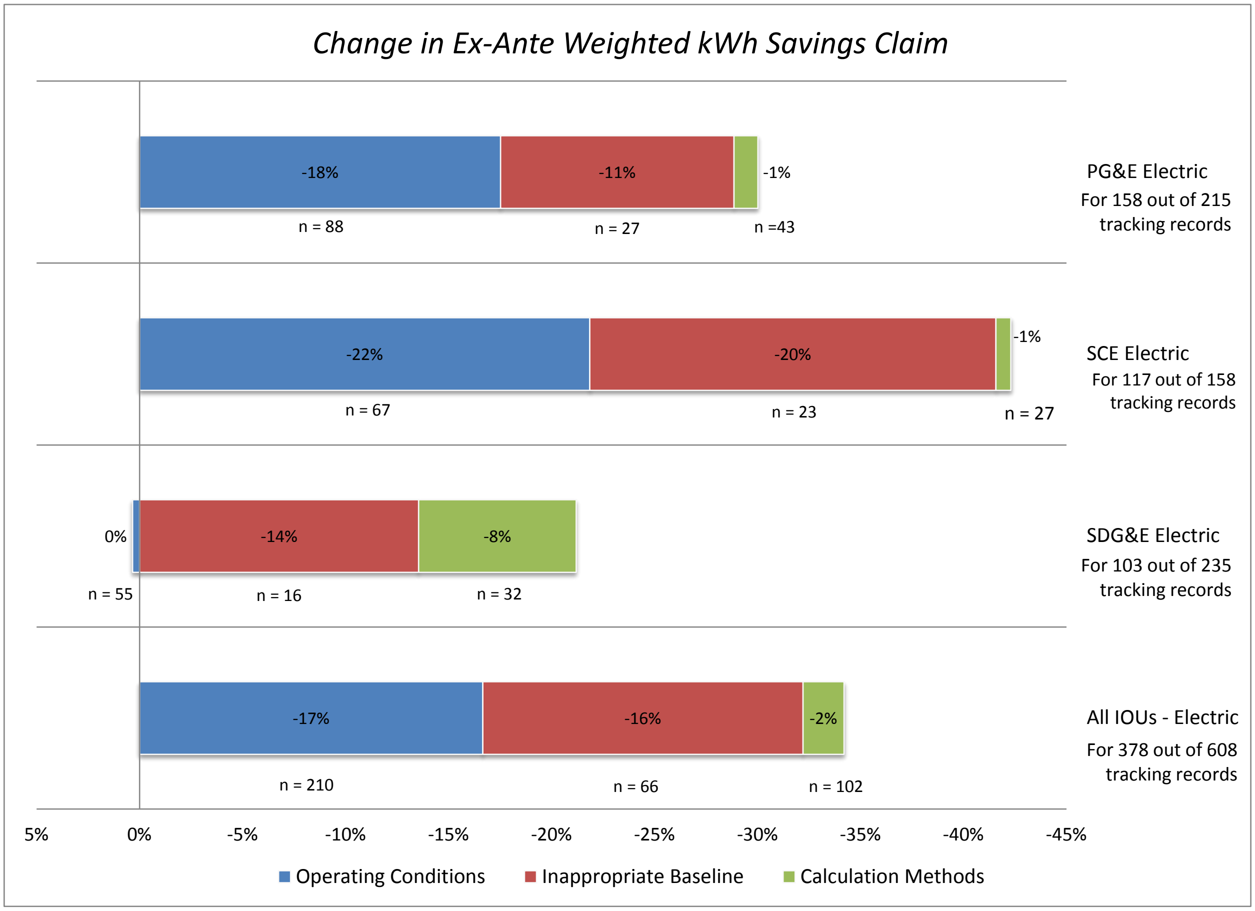
Figure 5-5: Frequency of Discrepancy Factors (All Projects and Records)



5.2.2 Effect on Ex-Ante kWh Savings Claims

Changes to ex-ante savings claims for the primary fuel assessed (kWh or therms) were calculated by record and summarized by IOU to show the energy impacts (in terms of percent) for the three largest discrepancy factors. Figure 5-6 details the domain-specific change in the ex-ante kWh savings claims for each of these three dominant factors.

Figure 5-6: Effect of Dominant Discrepancy Factors on Total Ex-Ante kWh Savings Claims (All Projects)

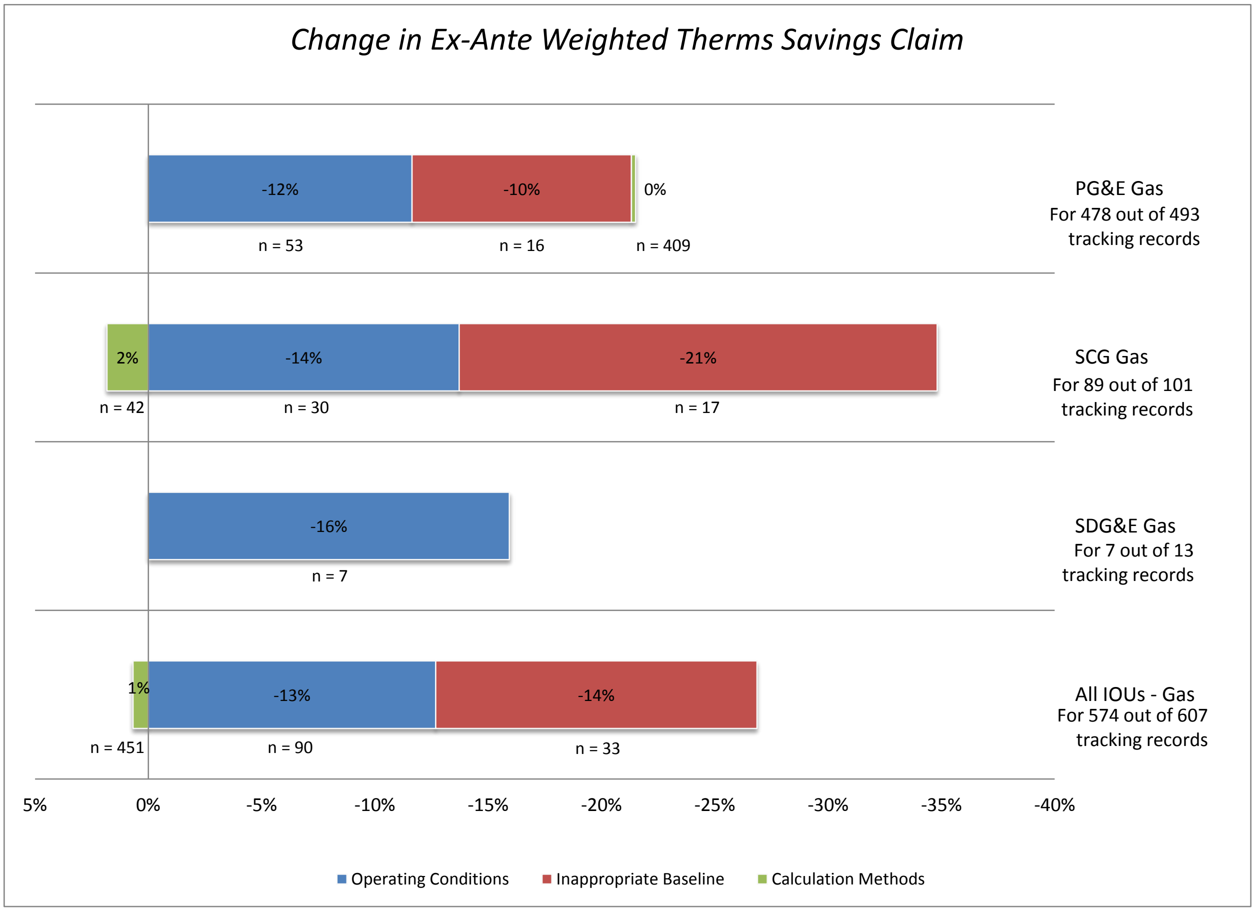


The gross impact results indicate that overall statewide ex-ante kWh savings were reduced by 35 percent for electric projects due to the three primary discrepancy factors described above. This accounts for the majority of the reduction in savings associated with the ex-post gross impact results. A total of 608 records distributed across 316 projects were analyzed. Within these electric projects, approximately 16 percent of this change resulted from baseline adjustments. Changes in operating conditions accounted for a 17 percent reduction of the total ex-ante kWh savings claims across IOUs, while changes in calculation methods resulted in a 2 percent reduction across total IOU kWh savings claims. Baseline issues were found to be a cause for discrepancy in 17 percent of the records analyzed, while operating conditions were found to be the cause for discrepancy for 45 percent of the records analyzed.

5.2.3 Effect on Ex- Ante Gas Savings Claims

Figure 5-7 summarizes the analysis of the domain-specific change in ex-ante gas savings claims for each of the three dominant discrepancy factors noted above.

Figure 5-7: Effect of Dominant Discrepancy Factors on Total Ex-Ante Gas Savings Claims (All Projects)*



*Excluding non-primary fuel projects, i.e., SDG&E electric projects that resulted in very high gas GRRs.

For natural gas projects, the total IOU ex-ante gas savings claim was reduced by 26 percent due to the three largest discrepancy factors. Approximately 13 percent of the reduction is due to changes from the operating conditions used in ex-ante savings estimation. Baseline adjustments for gas projects result in a 14 percent reduction in the total IOU ex-ante gas savings claim. Of the 1,215 tracking records, 607 are gas-savings measures associated with 179 gas projects. Nearly all gas tracking records were affected by these dominant discrepancy factors. However, the bulk of those observations are associated with the calculation method discrepancy factor, which yielded very little net change in savings. The operating condition and baseline specification factors, however, while occurring less frequently than calculation methods, both accounted for a considerable level of change to the ex-ante claims.

For both gas and electric projects, the findings suggest that IOU efforts to improve realization rates should focus on baseline selection and operating conditions, but with acknowledgment that project performance is affected dramatically for only a portion of the project records. For gas savings claims, 16 percent of cases were associated with operating conditions as the primary reason for discrepancy, and just 6 percent of projects associated with baseline selection.

Baseline issues are a larger contributor to differences between ex-post and ex-ante savings estimates for SCE electric and SCG gas projects, while baseline was not evident as an issue for SDG&E gas projects; note, however, that there were only 13 records analyzed for SDG&E gas projects. The use of different calculation methods resulted in a small change in ex-ante claims (+/- 3 percent or less) for all except one IOU fuel domain; the SDG&E electric domain shows that calculation methods are associated with an 8 percent decrease in ex-ante kWh savings claims in that domain, and nearly 19 percent of project records analyzed were found to have this primary issue.

While other IOU fuel domains show that a calculation method discrepancy occurred for about 20 percent of project records, 43 percent of SCG gas projects and 83 percent of PG&E's gas projects cite this discrepancy. Notable, however, is the relatively low net effect that this discrepancy has on ex-ante savings claims. This is because calculation method issues tend to improve savings in some cases while reducing them in others, and yield a limited effect overall. Baseline and operating conditions, on the other hand, yield a reduction in savings claims in most cases.

5.3 Program Level M&V Findings

This section presents gross impact realization rate results at the program level. Across all programs, a total of 429 points were sampled. Within the gross impact sample, 232 of the 429 points were sampled from core programs and 112 points were sampled from third party programs.[7] These two groups combined represent 80 percent of the sample points (344 out of 429 points). Table 5-4 presents GRRs by IOU fuel domain for the core and third party program groups. The average GRRs shown in Table 5-4 are unweighted and are representative of the sample but are not intended, or able, to be expanded to the subpopulations.

[7] A total of 33 SDG&E projects were submitted under the SDG&E Local Non-Residential (BID) program.

Table 5-4: Core and Third Party Program Group Gross Impact Results (Unweighted)

IOU Fuel Domain	Core Program				Third Party Implementation Program			
	N	GRR kWh	GRR kW	GRR Therms	N	GRR kWh	GRR kW	GRR Therms
PG&E Electric	51	0.75	0.69	n/a	43	0.49	0.52	n/a
PG&E Gas	39	n/a	n/a	0.54	19	n/a	n/a	0.80
SCE Electric	52	0.58	0.53	n/a	17	0.57	0.58	n/a
SDG&E Electric	28	0.60	0.61	n/a	29*	0.76	1.09	n/a
SCG & SDG&E Gas	62	n/a	n/a	0.67	4*	n/a	n/a	0.19*

* All SDG&E participation refers to projects submitted under the ‘Local03 - Local Non-Residential (BID)’ program

Table 5-5 presents GRR results for three additional program groups.

Table 5-5: NC, SGP, and LGP Program Group Impact Results (Unweighted)

IOU Fuel Domain	New Construction (NC) Program				Statewide Government Partnership (SGP) Program				Local Government Partnership (LGP) Program			
	N	GRR kWh	GRR kW	GRR Therms	N	GRR kWh	GRR kW	GRR Therms	N	GRR kWh	GRR kW	GRR Therms
PG&E Electric	5	0.84	0.62	n/a	6	0.59	1.03	n/a	7	0.38	0.26	n/a
PG&E Gas	4	n/a	n/a	0.99	10	n/a	n/a	1.86	3	n/a	n/a	0.37
SCE Electric	18	0.85	2.05	n/a	9	0.70	0.52	n/a	4	0.68	1.52	n/a
SDG&E Electric	14	0.47	0.63	n/a	2	1.14	1.79	n/a	0	-	-	n/a
SCG & SDG&E Gas	3	n/a	n/a	0.40	0	n/a	n/a	-	0	n/a	n/a	-

Table 5-4 and Table 5-5 highlight some differences between and within the IOUs’ core, third party, new construction, and government partnership programs. Due to a relatively small number of sample points for many program groups and IOU fuel domains and the fact that the results are unweighted, comparisons should be made cautiously. Relevant findings from this program analysis include:

- PG&E's electric core projects as a group yield a GRR of 0.75, one of the highest GRR results among the program groups presented. Of the 51 projects, one had a negative GRR and four had a GRR of zero. Approximately 45 percent of the projects had a GRR less than 0.50. Closer examination of programs that contribute to the core grouping shows very strong performance in the Agricultural and Pump Energy Efficiency Services programs, followed by the commercial program. The industrial program was a relatively low performer in this group; baseline and savings persistence were noted to be areas for attention, particularly in compressed air and wastewater treatment plant projects.
- PG&E's gas core projects have a GRR of 0.54, a relatively low result. However, this result is influenced heavily by one project with a GRR of -15.84. Projects with very high or very low realization rates had a substantial effect on average GRRs; the commercial core program has two projects with high GRRs in this group. Removing these projects, it appears that industrial projects are outperforming commercial projects in this core group. Furthermore, despite variation in mean GRR results, after considering the projects with extreme GRRs in an unweighted comparison, there may be no significant difference between the performance of PG&E's electric and gas core programs.
- For the PG&E electric domain third party program group, the average kWh GRR is 0.49. PGE2222 comprises 18 of the 43 projects in this group, and has a low average GRR that drives down the group result. One project had a negative GRR of -1.40 (E059) and two projects had GRRs (E009, E200) of 0.00. Inappropriate baseline and the evaluator's use of more current operating conditions were the primary issues that led to savings differences for the PG&E electric third party projects.
- The PG&E gas third party projects have a relatively high average GRR of 0.80. One project has a negative GRR and one project has a GRR of zero in this group, and 5 of the 19 projects have a GRR greater than 1.00. Three of three projects in the heavy industry program (PGE2223) have GRRs between 1.33 and 1.52. Three of five projects in a refinery program (PGE2225) have GRRs between 1.02 and 1.15. These are both larger third party programs. Four projects with a GRR below 0.50 (E089, E092, E244, E435) are from smaller programs (PGE2209, PGE2182, PGE2228, and PGE2234, respectively). *This finding points to a potential need for additional review of smaller third party projects and the training of third party implementers.*
- The new construction programs have relatively high average GRRs for both the PG&E electric and gas domains. However, there are a small number of sample points in each domain and project-level results are not consistently high. Of the five new construction electric projects in PGE21042, three are whole building projects with an average kWh GRR above one. However, two projects (E080, E236) remain challenged by low kW GRRs. Additionally, one VSD project (E236) has a very low kWh GRR of 0.03; the measure was found to be ineligible. The four projects in the PG&E gas domain for the new construction program (PGE21042) seem to indicate performance that is far above

average, but in fact only one project which involved new efficient gas griddles (E253) was found to have a realization rate close to 1.00. The GRRs for the two whole building projects in the gas domain are 0.05 and 0.28. The one HVAC project has a GRR of 2.74. *New construction programs are particularly in need of better kW estimation methods across all IOUs, but also should consider eligibility and baseline considerations. The results also point to the need to conduct thorough review of project applications and associated energy impact estimates, and understand all underlying savings assumptions and parameters for whole building projects. This is especially a concern for any projects using Energy Pro for ex-ante savings estimation. EnergyPro was designed as Title 24 compliance software, and embedded assumptions required for code compliance were found to be inappropriate for ex-ante savings estimates.* See Appendix F for more details on issues with Energy Pro modeling.

- SGP projects in the PG&E electric domain have an average kWh GRR and a high kW GRR based on a small sample. This group contains three MBCx (Monitoring-Based Commissioning projects) and one whole building project with low GRRs. Another whole building project performed much better, however. In the PG&E gas domain, the apparent over-performance for the ten sampled projects (an average GRR of 1.86), is driven by a whole building project with a GRR of 14.76. Only two of the ten projects have GRRs within 60 percent of unity, and seven projects have therm GRRs between 0.00 and 0.39.
- Results for SCE are fairly even across the program groups examined. For the core group, the kWh GRR ranges from 0.00 to 2.18, and the effect of projects with extreme GRRs – positive or negative - is dampened relative to other program examples mentioned throughout this subsection. The commercial core program was the strongest in the core group, followed by the agricultural program. The industrial program has a lower average GRR, as evidenced by four projects with zero savings. The commercial programs are fairly strong performers, and can be improved with better examination of HVAC and HVAC controls projects. The core industrial program could be improved by eliminating projects found to have a GRR of zero, through screening for eligibility (for example, the compressor in project F316 was not new, as required by program rules) and reconsidering the inclusion of smart wells (F029) as an eligible measure. In addition, monitoring for facility and store closures (F028) and following up with participants for persistence (such as in F001, where VSDs on well field pumps were not in use due to fracturing and well shutdown) could detect instances in which the measures failed to perform after relatively short periods and the ex-ante claim could be modified in the IOU's tracking system. *The agricultural projects -- comprising mostly pump rehabilitation projects -- need to better estimate peak demand savings and use the peak demand period definitions for energy efficiency programs*. Four pump rehabilitation projects had a kW GRR of zero, due to reporting the average demand savings during operation and not the peak demand savings during the CPUC DEER-defined peak periods per climate zone.

- The SCE third party programs have 17 projects with kWh GRRs ranging from 0.00 to 1.00. The group average kWh GRR (0.57) is lower than the domain average. Operating conditions is the most common discrepancy factor, occurring in seven projects. In almost all cases, the issue was that the operating hours were found to be less than those used in the original savings calculations. The three projects with a GRR of zero are distributed among these programs, and serve to significantly drive down the group GRR. Reasons for the GRRs of zero are measure disconnection due to the VSD causing motor burnout (F307), baseline issues for an insulation project (F428) and ISP requirements for high efficiency electric submersible oil well field pumps (F207). *Increased attention to baseline / ISP considerations and checks for the persistence of measures seems warranted for these third party programs; the measure in project F307 could have been removed from the IOU tracking database using periodic updates before final claims. The evaluation results also should be used by IOUs to identify when persistence is an issue for certain types of measures and to assist in determining measure eligibility. In addition, increased use of SCADA data or actual site operating logs before and after installation, as well as more conservative estimation of operating hours, could decrease discrepancies due to inaccurate operating hour estimates*.
- The SCE new construction program group appears to perform above average. This group has 18 projects with an overall kWh GRR of 0.85 (ranging from -0.11 to 5.11) and a kW GRR of 2.05 (ranging from -0.14 to 16.69). The program results are significantly impacted by two HVAC projects with inappropriate baseline (F216 and F357) that have very high kWh and kW GRRs. Without these two extreme projects, the kWh GRR for the new construction program group would have been 0.43 and the program kW GRR would have been 0.57. Pushing the kWh GRR lower are four projects with a GRR of zero; one project was found to use an improper baseline under Title 24 for cooling tower VSDs (F062) and another project involved the ISP for compressed air blowers at a wastewater treatment plant (F201). For two other projects, an HVAC and a whole building project, the buildings were not operational. *Increased attention after project start-up to actual operating conditions, and improved attention to baseline and ISP when claiming savings, would improve program GRRs for SCE's new construction programs*.
- The SCE Statewide Government Partnership program has nine projects with GRRs ranging from 0.10 to 1.08. The SCE Local Government Partnership program has only four projects ranging in kWh GRR from 0.44 to 0.95. In the SGP program group, two MBCx measures are low performers with a kWh GRR of 0.10 and 0.41. Baseline and measure eligibility issues exist for two HVAC measures. For the LGP group, the misalignment between kW and kWh GRRs resulted from one project, F027. The kWh GRR was 0.95 whereas the kW GRR was 4.37 for this project. This is a retro-commissioning project for which an incorrect calculation method was used to estimate peak demand savings. This led to a higher kW GRR of 4.36. Procedures used to generate ex-ante savings estimates for MBCx and retro-commissioning projects are areas for

improvement, as described in Appendix G. SCE should carefully examine HVAC and MBCx projects with respect to eligibility and baseline issues. Furthermore, SCE should use the DEER peak definition to estimate the peak demand savings of the energy efficiency projects.

- For the SCG and SDG&E gas domain, 69 projects were sampled. The core programs account for 62 of these projects. The therm GRR ranges from -0.75 to 2.60. The one project with a negative GRR (G418) is a boiler project with a regressive baseline (i.e., the installed equipment is less efficient than what was pre-existing and removed). There are nine projects with a GRR of zero, one for which a measure was not installed, three for which the equipment installed was not eligible, and four with baseline issues (primarily replacements of equipment at the end of its useful life). One project involved an eligible heat exchanger replacement that was shown not to save energy and the last project had operating condition changes that led to a zero GRR. SCG3607, the commercial core program, has 15 projects with a high group average GRR of 0.97, even with two boiler projects with a GRR of 0.00 and 0.43 (these projects had issues with eligibility and appropriate baseline). The 40 industrial core projects in SCG3611 have a lower group GRR of 0.54 as compared to the commercial core projects, despite having seven projects with a GRR between 1.02 and 1.49. The low GRR for the industrial core programs is due to the fact that all nine projects with a GRR of zero and the project with a negative GRR are in this program. Baseline and eligibility issues are the most important discrepancy factors for this program and should be a focus for improvement at all stages of project development. *SCG should, in its core program, apply more attention to correct project classification and ensuring that actual post-retrofit project conditions match those used in savings calculations*. Incorrectly classifying projects as early replacement continued as a common issue needing to be addressed, even in the well-established, utility–run core programs.
- Due to the small number of sample points, no firm observations can be made about performance in the gas domain for two relatively small programs - SDGE 3117 and 3118 (the Local Nonresidential BID program and the Nonresidential New Construction / Savings by Design program). The only third party program in the gross impact sample for SDG&E is SDGE3117. This program has a very small sample size of four gas projects and the therm GRR ranges from 0 to 0.41. An insulation project for SDG&E (H035) had the GRR of 0.05; the project involved replacement of worn insulation and was determined to be a repair and therefore ineligible, under program rules.
- Twenty-eight SDG&E core program group projects were evaluated, resulting in a mean kWh GRR of 0.60. The 22 core commercial projects in SDGE3105 have a group kWh GRR below the domain average, primarily due to six projects with a GRR of zero and one with a GRR of -0.04. Baseline issues emerged as dominant factors in three of these cases, with two cases of industry standard practice requirements and one VAV system

with Title 24 considerations. Three other projects involved telecommunications sites with upgraded power supplies. These telecommunication sites were found to be ineligible because the existing equipment was beyond its useful life and the new equipment was the only available replacement. Another site involved redundant claims for a project (H305); this project was claimed in two different programs within the evaluation sample. Eliminating these sites with zero savings would have raised the core program GRR to 0.77. This example highlights the need seen in other areas for more attention to project conformity with proper baseline and program rules. *SDG&E should ensure the measures being installed are not merely repairs or equipment replacement measures involving new equipment efficiency levels that are equivalent to existing equipment efficiency levels.*

- There are 29 third party program projects in SDGE3117 with an average kWh GRR of 0.76 and an average kW GRR of 1.09. The kWh GRR ranged from 0.00 to 3.71 and the kW GRR ranged from 0.00 to 9.16. There were four projects with very high kWh GRRs (above 2.00) in the group. Removing these four extreme points (consisting of one VSD and three refrigeration projects), the kWh GRR for third party programs drops to 0.45. Reasons for the low performance once these projects are removed vary across the five project types and are not obviously program specific issues. The 16 HVAC and HVAC controls projects have the lowest GRRs. Four projects with zero gross savings in the HVAC controls category have a significant impact on the results. H209 has zero savings because most of the energy saving features of the building controls system had been disabled due to operational problems. H421 is an HVAC control project determined to involve an ineligible repair measure. SDG&E should pay particular attention to HVAC projects in SDGE3117. More accurate inputs and a correct determination of useful life for the ice-building chiller in H307 and use of actual data as opposed to billing data for H004 would have improved project results and produced GRRs closer to unity.
- Fourteen SDG&E new construction program projects have an average kWh GRR of 0.47 and an average kW GRR of 0.63, substantially below the domain average for both metrics. Ten of the fourteen are whole building projects. Some whole building projects perform well with high GRRs, but many are below the domain average GRR. Two of the whole building projects have GRRs less than zero and three have a GRR of zero. Given the high variability and relatively low GRR for the group, closer attention to actual building operating hours and conditions, and appropriate use of the EnergyPro modeling tool for savings estimation would substantially improve program performance. *SDG&E should ensure that all whole building projects in the new construction programs utilize actual building operating conditions where possible and planned conditions when needed if actual conditions are unavailable.* Additionally, the EnergyPro model should be consistent with the physical as-built condition of the building. Equipment efficiencies and HVAC controls / sequencing should be used to estimate the ex-ante savings.

5.4 Measure Level M&V Findings

This section presents measure level findings by IOU fuel domain. The most common measure groups for each domain are presented, including unweighted realization rate statistics and the number of projects. Four measure groups were selected for presentation in each IOU fuel domain, with the exception of the PG&E electric domain, which supported five measure groups.[8] The related discrepancy factors associated with a given result are discussed for each measure group highlighted.

Some measure groups are common across domains, while some appear only in one domain. The measure groups are as follows:

- HVAC – includes chillers, air handling units, fans, and pumps
- HVAC Controls – includes sensors, control equipment, and EMS systems
- Variable Speed Drives – any variable speed drive controlling motors (for industrial or commercial use)
- Compressed Air Systems – compressors, ancillary components in compressed air systems, and controls
- Heat Recovery – water side heat exchangers and economizers
- Boilers - new or upgraded boilers and their ancillary systems; boiler controls for industrial applications
- Electronically Commutated (EC) motors – EC motors and controls for refrigeration system evaporators (both refrigerated cases and walk in coolers / freezers)
- Pump Rehabilitation – pump overhaul and improvements for agricultural customers and water distribution agencies
- Steam Traps – steam trap replacements in high and low pressure service
- Whole Building / Envelope – comprehensive packages of energy efficiency measures, typically in new construction projects that utilize building simulation models such as Energy Pro.
- Refrigeration – new or upgraded refrigeration systems and associated controls

5.4.1 Measure Level Findings for the PG&E Electric Domain

In the PG&E electric domain, there are a total of 112 projects. The five most common measure groups representing 65 of the 112 projects and 52 percent of the domain savings are summarized in Table 5-6.

[8] The PG&E electric fuel domain also possesses the largest number of sample points when compared to other domains.

Table 5-6: Measure Level Unweighted Impact Results for the PG&E Electric Domain

PG&E Electric Domain								
Variable Speed Drives (VSD) (N=17)			**HVAC Controls (N=16)**			**Electronically Commutated (EC) Motors (N=13)**		
Mean	RR - kWh	RR - kW	**Mean**	RR - kWh	RR - kW	**Mean**	RR - kWh	RR - kW
	0.90	**0.76**		**0.49**	**0.42**		**0.57**	**0.67**
Site ID	**RR - kWh**	**RR - kW**	**Site ID**	**RR - kWh**	**RR - kW**	**Site ID**	**RR - kWh**	**RR - kW**
E204	6.39	5.30	E124	2.69		E337	1.16	1.40
E093	2.54	2.01	E201	1.22	3.08	E338	0.99	1.10
E079	1.55	0.89	E319	1.12	1.01	E341	0.82	0.91
E217	1.03	1.03	E227	1.03		E340	0.79	0.91
E052	0.92	0.92	E106	0.84	0.94	E235	0.68	0.83
E325	0.85	0.85	E322	0.72		E111	0.53	0.60
E233	0.65	0.11	E301	0.55	0.75	E122	0.51	
E218	0.49	0.49	E308	0.45	0.00	E107	0.44	0.50
E202	0.42	0.43	E081	0.39	0.39	E232	0.42	0.51
E419	0.38		E306	0.23	0.28	E116	0.41	0.50
E224	0.35	0.29	E312	0.20	0.13	E342	0.38	0.44
E211	0.35	0.34	E304	0.11	-0.05	E109	0.31	0.38
E065	0.25	0.32	E209	0.09		E234	0.00	0.00
E205	0.23	0.23	E210	0.08				
E402	0.17	0.17	E084	0.00				
E302	0.09	0.17	E426	-1.87	-2.29			
E059	-1.40	-1.40						

HVAC (N=12)			**Compressed Air (N=7)**		
Mean	RR - kWh	RR - kW	**Mean**	RR - kWh	RR - kW
	0.68	**0.43**		**0.43**	**0.51**
Site ID	**RR - kWh**	**RR - kW**	**Site ID**	**RR - kWh**	**RR - kW**
E339	2.63	1.08	E408	1.12	1.12
E046	0.99	1.08	E305	1.10	1.12
E303	0.90	-0.57	E323	0.48	0.49
E208	0.87	0.50	E321	0.28	0.29
E040	0.87	-0.13	E041	0.01	0.01
E005	0.76	0.65	E057	0.00	0.00
E010	0.44	0.44	E123	0.00	
E072	0.30	0.72			
E037	0.24	0.90			
E228	0.08				
E236	0.03	0.10			
E076	0.00	0.00			

Measures groups in the PG&E electric domain have mixed results compared to the domain averages of 0.67 for kWh and 0.52 for kW. The VSD and HVAC groups seem to perform well with high GRRs for kWh. The HVAC controls and compressed air groups perform less favorably, and the EC motors group results are average. Due to the presence of projects with extreme GRRs, however, performance for the VSD group may not be as high as indicated by the mean GRR alone; conversely, performance in the HVAC controls group may be not as poor as indicated by the mean GRR.

- For the VSD group, one data center project, E204, with a kWh GRR of 6.39 and a kW GRR of 5.30 drives the relatively high group kWh GRR of 0.90. Without this point, the mean kWh GRR is 0.55. On the other hand, project E059 has a kWh GRR of -1.40. Removing that point (prior to other adjustments) adjusts the overall VSD GRR up to 1.04 from 0.90. Both projects - E204 and E059 - are associated with baseline discrepancy factors and illustrate the importance of appropriate baseline estimation.

- Also for VSDs, there are a relatively large number of projects, 13 out of 17, affected by the operating conditions discrepancy factor. In most instances, these operational updates led to a downward adjustment to ex-post savings relative to ex-ante claims. The IOU calculation methods, whether spreadsheet or software programs, appear appropriate, but often operating hours were overstated or actual motor loading and its variation was not accounted for. Collection of spot and trend data from the participants' control and operating systems, along with the application of that data to savings, should be considered. *IOUs can improve results by carefully reviewing inputs, collecting more current inputs, and updating the savings calculations for VSDs.*

- In the HVAC controls group, mean GRRs are below 0.50, and range from -1.87 to 2.69. This suggests that ex-ante impacts are particularly uncertain for this measure offering. Project E426 has the negative kWh GRR of -1.87; removing this one record, the mean result for this measure is 0.65 instead of 0.49, approaching the PG&E electric domain mean. *The wide range of GRRs and the ex-ante conditions suggest possible improvements in savings estimation techniques and point to better data collection for this measure group. Data collection should center on the most important variables on which savings were based (e.g., temperatures, hours of use, etc.) post-retrofit and after systems have stabilized into their new mode of operation. It should be clearly identify when strategies have been discontinued or control sequences have changed from those that were planned, as not all control changes function exactly as intended and some changes have unintended consequences.*

- There are a total of 13 EC motor projects, 9 discrepancies are associated with calculation methods and 2 due to measure count discrepancies. GRR results for this measure are average, without extremely high or low performing projects. GRRs are below 0.50 for 6 out of 13 projects and only exceeded 1.00 in one instance; one vacant facility had zero savings. The ex-post impacts were derived using different calculation procedures in most instances. The evaluation team used custom spreadsheets based on IOU workpapers and adjusted for climate zones.

- For the HVAC group, the high average kWh GRR of 0.68 is noted, however, a project with a GRR of 2.63 is one cause for this result. On the other hand, projects E076 showed baseline issues and E236 showed calculation method changes and inoperable measures as the primary reasons for low project kWh GRRs of 0.00 and 0.03 respectively, reducing group GRR. Ex-post savings that have a considerable variance in either direction from the ex-ante savings are a cause for concern.

- Also, in the HVAC group, peak demand (kW) energy metric results are somewhat low, due in part to two negative realization rates for projects E040 and E303. One negative GRR is due principally to calculation methods, while the other is due to baseline selection and operating conditions. Inappropriate baseline selection in the ex-ante models is an issue for two data center HVAC projects (E208 for a VSD on chiller, E303 waterside economizer system). These projects have relatively high kWh GRRs (0.87 and 0.90) but poor kW GRRs (0.50 and -0.57). *The HVAC measure group calculation methods for estimation of peak demand should be examined; it should be verified that the power consumption at peak and non-peak loads are differentiated and that the kW savings claimed are during the DEER-defined peak periods for that climate zone. Baseline issues can often* by thoroughly reviewing appropriate baseline documents and selecting appropriate HVAC equipment technology, equipment sizing, controls, and set points while building the baseline model for the ex-ante estimate. Operating conditions discrepancy can be minimized by incorporating the actual as-built operating profile to the savings estimate during the IR application process.

- For the seven compressed air projects, group GRR results are low, and do not include extremely high or extremely low performing projects. However, GRRs of zero or near-zero are common for this measure (three out of seven projects), and are due to baseline issues in two projects and operating conditions in one other project. Two projects both had GRRs in excess of one, with discrepancies due to calculation methods. PG&E used a variety of calculation methods, while the evaluators used the DOE tool *(AirMaster)* in most cases. Operating hours account for much of the discrepancy in compressed air projects. Equipment operation changes (and not normalizing for actual airflow and pressures) were frequent reasons for low GRRs in these projects. *It is recommended that PG&E attempt to better capture post-retrofit operating conditions and adjust savings, as well as adopt more robust modeling approaches like AirMaster for compressed air systems.*

5.4.2 Measure Level Findings for PG&E Gas Fuel Domain Projects

In the PG&E Gas domain, there are a total of 75 projects, with 33 projects occurring within the largest four measure groups identified in Table 5-7. The 34 projects account for 24 percent of the domain savings.

Table 5-7: Measure Level Unweighted Impact Results for the PG&E Gas Domain

PG&E Gas Domain			
Heat Recovery (N=12)		**Boilers (N=7)**	
Mean	RR - Therms	**Mean**	RR - Therms
	0.75		**0.58**
Site ID	**RR - Therms**	**Site ID**	**RR - Therms**
E245	1.52	E350	1.09
E430	1.12	E001	0.98
E002	0.99	E336	0.79
E352	0.94	E098	0.67
E003	0.91	E092	0.49
E105	0.66	E048	0.03
E242	0.72	E244	0.00
E243	0.93		
E240	0.95		
E435	0.27		
E014	0.00		
E246	0.00		
HVAC Controls (N=8)		**HVAC (N=6)**	
Mean	RR - Therms	**Mean**	RR - Therms
	2.33		**-1.97**
Site ID	**RR - Therms**	**Site ID**	**RR - Therms**
E121	9.72	E255	2.74
E113	5.23	E064	1.33
E115	2.19	E333	0.00
E013	1.50	E444	0.00
E332	0.77	E443	-0.06
E119	0.16	E334	-15.84
E441	0.00		
E118	-0.93		

Measure groups in the PG&E gas domain have mixed results compared to the domain average of 0.67 for therms. The boilers and heat recovery groups show GRRs within ten percent of the average domain GRR, but the HVAC group shows very high mean GRRs and the HVAC group shows extremely low negative GRR.

- In the heat recovery group, there are a relatively large number of projects, 8 out of 12, affected by the operating conditions discrepancy factor. There are two projects with zero savings; the overall GRR without these two projects would be relatively high, at 0.90 (close to the IOU default GRR). In E014, for instance, one of the projects with a GRR of zero, the evaluators found that the change to surge drum operation in the refinery's process was not successful on a long term basis, but was effective at the time of the IOU's post-installation site visit. In one project (E245), the IOU-claimed baseline condition of early replacement was found to be normal replacement. One heat exchanger (project E246) was not found to be above standard practice and had a GRR of zero. *As with other sampled projects, IOUs should be vigilant about ensuring likely measure persistence while approving a project, using conservative assumptions for expected performance and consider flagging projects for additional post installation activities if operation is not stable.*

- The HVAC controls category as a group had a very high GRR of 2.33. However, HVAC control measure GRRs range from -0.93 to 9.72. The wide range shows the limited precision and accuracy of ex-ante savings estimates. For projects E115, E118, E119 and E121, the ex-ante baseline models lacked proper calibration and the inputs for onsite conditions (including set points and temperatures) were not properly captured and inappropriate. Project E441 has a GRR of zero as the project did not exceed Title 24 baseline requirements. *PG&E should expand detailed data collection (including building characteristics, HVAC fan schedules, temperature set points, night setback, operating schedules, equipment loading and control sequences) during ex-ante baseline model development, and use meter or interval data to calibrate models when they are used. Additionally, the models should be thoroughly reviewed.*

- The boiler group shows relatively good results. However, baseline or measure eligibility issues were each found in two of the seven measures (a total of 29 percent) of the projects in this group. Project E092 and E244 were determined to be ineligible, as the measure was an operational change; E092 involved shifting boiler load to an existing more efficient boiler. In one project (E098), the IOU-claimed baseline condition of early replacement was found to be normal replacement.[9] For E048, a low GRR of 0.03

[9] Examples like this project, which incorrectly classified a normal replacement project as an early replacement project, highlight a frequently observed problem.

resulted from a large overestimation of boiler loading. *It is recommended that PG&E conduct in-depth reviews to verify measure eligibility per program rules and select appropriate baseline conditions to further improve GRR results for boiler measures. Additionally, careful review of applications and final savings are needed to prevent large discrepancies in energy savings.*

- Given the extremely low group GRR of -1.97, HVAC measures appear underperforming, but sensitivity analysis indicates that just one project, E334, with a GRR of -15.84, plays the greatest role in driving down the GRR result. Without this point, the group GRR is 0.80, within ten percent of the domain GRR. The range of HVAC measures GRR was from -15.84 to 2.74. For project E334, an operating conditions change and measure counts caused most of the discrepancy; the baseline model was built with incorrect information related to scheduling (the units remained off during unoccupied hours in the pre-retrofit period and no energy savings result from the temperature setback measure). For E333 and E443, the baseline was incorrect; existing equipment was past its useful life and code required equipment was used instead of existing in situ equipment. In project E255, with a therm GRR of 2.74 and an electric GRR of 0.60, the ex-ante Energy Pro model did not capture the savings found in the ex-post analysis. *PG&E should more carefully identify the remaining useful life of equipment and use code requirements to determine proper baselines when normal replacement baselines apply. This would serve to improve savings estimation accuracy over the measure life and reduce the wide GRR range. Furthermore care should be taken to capture actual operating parameters and selecting the most appropriate models for gas HVAC projects.*

5.4.3 Measure Level Findings for SCE Electric Domain Projects

In the SCE electric fuel domain, there are a total of 98 projects. Of those, 56 projects are in the four most common measure groups identified in Table 5-8; these projects represent 70 percent of the domain savings.

Table 5-8: Measure Level Unweighted Impact Results for the SCE Electric Domain

SCE Electric Domain					
Variable Speed Drives (VSD) (N=16)			**Water Pump Rehabilitation (N=15)**		
Mean	RR - kWh	RR - kW	**Mean**	RR - kWh	RR - kW
	0.64	**0.79**		**0.62**	**0.30**
Site ID	**RR - kWh**	**RR - kW**	**Site ID**	**RR - kWh**	**RR - kW**
F011	1.05	3.19	F324	1.47	0.00
F228	1.03	1.06	F325	1.10	-0.01
F317	1.00	0.84	F059	1.06	1.74
F002	0.97	0.99	F315	1.06	0.14
F051	0.96	0.59	F069	1.01	1.06
F301	0.95	0.99	F326	0.92	0.00
F009	0.73	0.86	F066	0.76	0.00
F313	0.64	0.91	F423	0.44	0.38
F005	0.58	0.59	F422	0.40	0.23
F041	0.52	0.66	F042	0.38	
F015	0.52	0.56	F008	0.26	0.23
F007	0.41	0.42	F223	0.17	0.22
F035	0.35	0.46	F030	0.15	0.16
F012	0.34	0.23	F053	0.12	0.09
F004	0.26	0.26	F061	0.00	0.00
F001	0.00	0.00			
HVAC (N=14)			**Compressed Air (N=14)**		
Mean	RR - kWh	RR - kW	**Mean**	RR - kWh	RR - kW
	0.99	**2.58**		**0.45**	**0.53**
Site ID	**RR - kWh**	**RR - kW**	**Site ID**	**RR - kWh**	**RR - kW**
F216	5.11	16.69	F306	1.21	1.37
F357	3.41	11.12	F050	1.00	1.00
F203	1.09	0.78	F018	0.80	0.83
F314	0.92		F215	0.74	0.70
F032	0.87	2.26	F026	0.58	0.58
F219	0.70	0.70	F303	0.48	0.81
F013	0.66	0.37	F208	0.46	0.94
F302	0.51	0.74	F430	0.36	0.36
F412	0.35	0.69	F057	0.29	0.37
F202	0.25	0.18	F220	0.16	0.16
F023	0.09	0.13	F024	0.16	0.14
F429	0.00	0.00	F022	0.11	0.11
F058	0.00	0.00	F028	0.00	0.00
F200	-0.11	-0.14	F316	0.00	0.00

Measures groups in the SCE electric domain have a mixed performance compared to the domain average GRRs of 0.61 for kWh and 0.57 for kW. Average group GRRs for energy (kWh) are close to the domain mean for only two measures; the group kW mean averages are very widely dispersed from the domain kW mean. The HVAC group seems to perform very well with a high GRR of 0.99 for kWh. The compressed air group performs poorly in both kW and kWh results.

- For VSDs in the SCE electric domain, as with the PG&E electric domain, a relatively large number of projects, 10 out of 16, were affected by the operating conditions discrepancy factor. In most instances, these operational updates led to a downward adjustment to ex-post savings relative to ex-ante claims. In a number of cases, the IOUs used forecasting to estimate measure operation and determine savings from VSDs over a five year period. Avoiding the use of forecasting models that use projected flow rates, throughput, etc., will increase the accuracy of ex-ante savings estimates. Results were negatively impacted by F001, a large stratum 1 project involving a VSD on an underground oil well pump; this project has a zero GRR because the equipment was non-operational due to a change in well condition experienced after installation. *Realization rates could be improved for this measure group through increased attention to ex-ante project descriptions for accurate baseline determination and more conservative assumptions to better address operating conditions, especially for large projects. To address persistence, all IOUs should periodically review the performance of large projects months and years after installation, not only to mitigate risks to energy savings claimed but also to better serve program participants and expand energy saving potential.*

- The predominant cause for discrepancy factors in pump rehabilitation projects was operating conditions; in ten of the fifteen projects, post-installation operating conditions, such as hours of use and water production, were different than ex-ante estimates. Four pump rehabilitation projects had a kW GRR of 0.00, due to reporting the average and not the peak demand savings (the evaluators were better able to determine when demand savings occurred through operator interviews and with pre-retrofit and post-retrofit interval data in many cases; the IOUs could also use a smaller set of interval data to estimate when pump savings are likely to occur). *Where appropriate, all IOUs should use smart meter and interval data to estimate energy and demand savings for this and other measures; spot measurements and production data should be used to verify these estimates.* Interval records are especially valuable to determine peak demand savings. More accurate reporting and continuous monitoring of site operation would also significantly improve GRRs for the measure.

- Although the average GRR of 0.99 for the HVAC group appears nearly optimal, this was mostly due to two specific projects with individually high kWh GRRs. F216 has a GRR of 5.11 due to inaccurate design load estimates; F357 had a GRR of 3.41 due to an inappropriate baseline selection and changes in operating conditions. Without these two outliers, the average GRR for HVAC would have been 0.45, below the average domain GRR. The kW GRR would be reduced to 0.52 from 2.58. Changes in operating conditions (observed in 6 of 14 projects) and baseline specification (observed in 5 of 14 projects) are the primary reasons for low project GRRs. Baseline and eligibility remain issues for SCE. For projects F013 and F032, the ex-ante baseline model lacked proper calibration and the inputs for onsite conditions were not properly captured. F302 involved an ineligible VAV installation. As recommended for PG&E, SCE should also employ more thorough data collection *(including set points, temperatures, schedules and control sequences) during ex-ante baseline model development, and use meter or interval data to calibrate models when they are used.* Additionally, the models should be thoroughly reviewed.

- Compressed air projects as a group had low performance. Low GRRs (two projects with zero savings and three with a GRR less than 0.16) seriously impact results. Operating conditions, calculation methods, and ineligible measures account for most of the discrepancies found. As an example of calculation method deficiencies, in projects F024and F220, the ex-ante calculations failed to normalize savings to annual compressed air usage. Three of the fourteen compressed air projects were ineligible. Leak repair measures were not eligible for rebates during some periods of this evaluation cycle, and their savings were therefore removed from the project gross impacts. In project F316, the air compressor was not new when installed, but was instead relocated from another plant prior to the project implementation (in violation of program rules). For F306, SCE claimed a replace on burnout baseline; the evaluators, based on the on-site visit, assessed an early retirement baseline for the existing gas-fired wastewater treatment plant process air compressors; the change to a new electric-driven compressor resulted in project level kWh GRR of 1.21. Lower as-found hours of use were the primary cause for discrepancy for F057, F208 and F215. To address these deficiencies, SCE should review pre-retrofit and post-retrofit air flow and normalize savings to represent typical annual operating airflow consumption for compressed air projects. *Closer attention to more accurate calculation methodologies and inputs, accounting for hours of use and normalizing for airflow for production effects, would have substantial impact on the measure category and raise overall group realization rates*.

5.4.4 Measure Level Findings for SCG and SDG&E Gas Fuel Domain Projects

In the combined gas fuel domain for SCG and SDG&E projects, there are a total of 69 projects. The measure groups with the highest number of projects are shown in Table 5-9 below, accounting for 35 projects and 61 percent of the domain savings.

Table 5-9: Measure Level Unweighted Impact Results for the SCG / SDG&E Gas Fuel Domain

SCG and SDG&E Gas Domain			
Boilers (N=13)		**Heat Recovery (N=12)**	
Mean	RR - Therms	**Mean**	RR - Therms
	0.68		**0.76**
Site ID	**RR - Therms**	**Site ID**	**RR - Therms**
G207	1.71	G030	2.12
G218	1.31	G402	1.24
G011	1.16	G002	1.22
G404	1.04	G003	0.74
G401	1.02	G407	0.69
G403	0.97	H001	0.67
G022	0.86	G405	0.57
G309	0.69	G026	0.56
H427	0.00	G211	0.52
G221	0.44	H006	0.41
G206	0.43	G001	0.33
G013	0.00	G021	0.00
G418	-0.75		
Steam Traps (N=5)		**HVAC Controls (N=5)**	
Mean	RR - Therms	**Mean**	RR - Therms
	0.96		**0.47**
Site ID	**RR - Therms**	**Site ID**	**RR - Therms**
G307	2.60	G040	0.91
G419	1.06	G204	0.76
G305	0.85	G219	0.39
G318	0.28	H010	0.30
G313	0.00	G024	0.00

Measures groups in the SCG and SDG&E gas domain have fairly good results compared to the domain averages of 0.58 for therms (this domain result excludes the projects with the most extreme GRRs). Steam traps as a group seems to perform very well, but the results vary widely over a range of zero to 2.60. The boiler and heat recovery group show GRRs 10 percent greater than the average domain GRR, and the HVAC controls group shows a group mean that is 11 percent lower than the domain mean.

- A total of 13 boiler projects are affected by the following discrepancy factors: four with changed operating conditions, four with changes to calculation methods, two with inappropriate baselines and ineligible measure, and one each for equipment specifications and unquantified fuel impacts. Two individual projects are extremes with GRRs of 1.71 and -0.75. G418 is a normal replacement project with a GRR of -0.75 that installed new equipment with an efficiency rating below that of the removed equipment, resulting in negative savings associated with inappropriate ex-ante baseline specification. In G207, the boilers were still operating, in good condition, and able to meet facility needs; the baseline was changed from standard practice to the pre-existing equipment, resulting in a significant increase in savings. Two projects (G013 and H427) have zero savings and reduce the mean GRR. SCG and SDG&E should clearly understand the baseline requirements for boilers and other technologies in order to prevent over-estimating or under-estimating project savings. As operating conditions and calculation methods are cited primary discrepancy factors, SCG and SDG&E need to more closely examine actual ex-post operating conditions and scrutinize calculation methods for boilers.

- Half of the 12 projects in the heat recovery group are affected by the operating conditions discrepancy factor. In four of the six projects with this primary discrepancy, production rates were different than those used in ex-ante calculations. The results were significantly affected by one project (G030) with a relatively high GRR of 2.12, which resulted from a misinterpretation of site supplied production data. This was offset by a project with a GRR of 0.00. Based on this relatively small sample of projects, this area follows the trend of other categories with improvements in assessing actual operating conditions - including production data for heat recovery measures - being an important opportunity area which could raise overall GRR.

- The steam trap group has only five projects. However, two projects in the group clearly demonstrate the importance of calculation methods for steam traps and the need for consistent evaluation methodology. One project, G307, had a large GRR of 2.60. This project used a deemed approach to calculate savings instead of the more appropriate custom methodology for industrial steam traps. This project raised the overall average for the measure category to 0.96 from 0.55. Another project, G318, also used the deemed approach to calculate savings. However, in this case, accurate accounting for the operating pressures and orifice sizes resulted in a low GRR of 0.28. Savings were also lowered by an ineligible project; project G313 was ineligible because known blocked or failed closed steam traps are not eligible. SCG should standardize on steam trap calculation approaches in the program rules and review projects for both the proper calculations and inputs, as well as for eligibility criteria.

- The HVAC controls measure group had only five projects. From this small sample, baseline appears to be a dominant issue. Project G024, with GRR of zero, involved an original system being replaced at the end of its useful life with equipment not exceeding Title 24 requirements. The group average GRR would have been 0.59 without the zero savings project. Project G219 also had a baseline issue in addition to operating condition changes which resulted in a low project GRR of 0.39. For these projects, we can see common issues concerning baseline, and proper project screening and application review.

5.4.5 Measure Level Findings for SDG&E Electric Fuel Domain Projects

In the SDG&E Electric fuel domain, there are a total of 73 projects; 50 projects are within the top four measures groups, as shown in Table 5-10. These projects account for 80 percent of the domain savings.

Table 5-10: Measure Level Unweighted Impact Results for the SDG&E Electric Domain

SDG&E Electric Domain					
HVAC (N=16)			**HVAC Controls (N=12)**		
Mean	RR - kWh	RR - kW	**Mean**	RR - kWh	RR - kW
	0.36	**0.50**		**0.52**	**0.60**
Site ID	**RR - kWh**	**RR - kW**	**Site ID**	**RR - kWh**	**RR - kW**
H009	1.29	2.26	H214	1.12	1.13
H020	1.06		H218	0.91	
H014	0.59	1.08	H321	0.77	0.48
H025	0.45	0.00	H042	0.61	0.70
H027	0.41	0.31	H028	0.38	
H423	0.37	0.30	H205	0.13	
H024	0.35	0.69	H402	1.21	
H302	0.32	0.23	H202	1.15	2.45
H004	0.20	1.54	H013	0.00	0.00
H026	0.19	0.00	H029	0.00	0.00
H201	0.14	0.13	H209	0.00	0.00
H007	0.14	0.13	H421	0.00	0.00
H307	0.11	0.18			
H206	0.11	0.19			
H208	0.00				
H305	0.00	0.00			
Whole Building / Envelope (N=11)			**Refrigeration (N=11)**		
Mean	RR - kWh	RR - kW	**Mean**	RR - kWh	RR - kW
	0.48	**0.79**		**1.46**	**1.73**
Site ID	**RR - kWh**	**RR - kW**	**Site ID**	**RR - kWh**	**RR - kW**
H031	1.14	1.28	H322	3.71	0.00
H416	1.11	2.32	H016	2.79	0.62
H200	1.03	1.44	H051	2.40	2.40
H032	0.96		H023	2.29	9.16
H034	0.61		H005	1.48	1.55
H317	0.57	1.74	H002	1.42	1.92
H401	0.47	0.51	H406	0.97	0.89
H046	0.07	-0.11	H045	0.73	0.73
H008	0.00	0.00	H044	0.30	0.00
H311	-0.14	-0.21	H040	0.00	
H308	-0.54	0.11	H203	0.00	0.00

Three of the four most common measure groups in the SDG&E electric domain have lower performance compared to the domain averages of 0.64 for kWh and 0.82 for kW. The refrigeration group has higher performance, driven somewhat by projects with very high realization rates.

- For the HVAC group, most projects performed poorly; 13 of 16 projects have a kWh GRR below 0.50. Only two projects have kWh GRRs greater than one. There are two projects that had GRRs of zero. Several discrepancy factors have a large negative impact on the group GRR. Ineligible measures drive down the GRRs in this group. Inappropriate baseline also brought the average GRR scores for the HVAC group down considerably; the average GRR for those three projects was 0.07 for kWh and 0.14 for kW. For H208, the installed air handling units were the same as required by Title 24. The other two projects (H206/H307) were the same project with two partial payments and involved a thirty year old chiller well past its useful life. Tracking database discrepancies occurred for two projects. Calculation methods and savings estimates for this group employ billing analysis, deemed methods and other techniques that can, if not appropriate for the specific measure and applied correctly, result in inaccurate savings. For one project (H009), the ex-ante savings was based on billing analysis and substantially underestimated savings. H004 used billing analysis and had a very low result for kWh (0.20) but a high kW result (1.54 GRR). Calculation method changes resulted in a low kWh GRR (0.19) for H026. *To correct these deficiencies, SDG&E should use robust savings estimation techniques and improve their calculation methods, including site-specific spreadsheet analysis with appropriate input values for HVAC measures, in order to accurately determine energy savings and avoid underestimating or overestimating savings. In addition, SDG&E should carefully review electric projects for eligibility, adherence to program rules, proper baseline determination, and accurate tracking system data.*

- The HVAC controls group also performs poorly. While three projects had kWh or kW GRRs higher than one, there were four projects that had GRRs of 0 – two ineligible measures (repairing or replacing like for like as in project H421, in which the measure was not eligible because it entailed repairs to CO sensors), one program rule violation, and one inoperable measure (in H209, where pressure set point strategies were not implemented). The ex-post evaluation found six projects in this group had kWh GRRs less than 0.38. Calculation methods led to a low realization rate for one particular project; a low GRR was found for H205 and was primarily due to the differences in modeling guestroom fan coil unit operation, set points, and schedules. *SDG&E should carefully check program eligibility and ensure that repair or replacement projects are not included. SDG&E should carefully confirm that projects are clearly identified to assist in screening these projects. Operating conditions should be checked after* *installation and savings updated. Further, SDG&E needs to check that all fuel impacts are considered.*

- For the whole buildings group, seven of eleven projects list changed operating conditions as the primary discrepancy. There are two projects with negative kWh GRRs (H311 and H308). Unquantified fuel impacts for one large strata 1 project (H008) resulted in zero savings. This project has a cogeneration plant on-site and most of the electricity needs of this project are met by the electricity generated from the cogeneration plant. Hence, the electricity generated by the cogeneration plant was not eligible. The energy savings claimable for this project are therefore limited to the amount of electricity provided by SDG&E to the site. Given the wide spread of results, *SDG&E should make certain that opening conditions are as accurately represented as possible, that savings are adjusted in project documentation and tracking systems, that appropriate models are used, and that tracking systems contain the correct savings values in the whole building group. Further, SDG&E and other IOUs should study projects that perform well in terms of saving energy and have good evaluation results, in order to replicate those project successes*. As a final note, the secondary fuel gas result (GRR=10.15) was very high, driven by a few outliers. SDG&E should devote special attention to correctly calculating any secondary fuel impacts.

- The refrigeration group had kW and kWh GRRs substantially above 1.00. Six of eleven projects had kWh GRRs substantially greater than 1.00. Besides two ineligible projects, the only project with a poor kWh GRR (0.30) was H044; this was related to lower energy savings at the compressors from the case night covers at this supermarket. For this project, the energy savings occur only at night, and the peak demand savings were thus found to be zero (the ex-ante calculations used average demand savings over the entire year, including daytime hours). Use of deemed approaches led to a high kWh GRR (3.71) and a zero kW GRR for H322. Of note, one very large kW GRR (9.16) was found for H023. While realization rates close to or above one are appropriate, very high realization rates point to the need for SDG&E to examine calculation approaches and to capture true operating conditions and peak demand savings for this measure type. *The IOUs should also avoid the use of deemed savings where not appropriate.*

5.4.6 Summary of Measure Level Findings

Findings arising from the measure level analysis are summarized in this subsection. The findings are instructive in comparing results and issues across measure groups. For example, do certain measure groups perform consistently better (with higher realization rates) than other groups? Do any groups have particularly pronounced reasons for discrepancies between ex-ante and ex-post savings? Which groups might benefit from specific attention with regard to program design, project development, and project installation and verification? The findings are as follows:

- The VSD measure group was analyzed for both the PG&E electric and the SCE electric domains. This group has a high GRR for PG&E and an average GRR for SCE. However, the high GRR for PG&E results from one extreme sample point. Operating conditions is a common reason for discrepancy, as loads (as well as hours) varied from ex-ante estimates. In a number of cases, the IOUs used forecasting to estimate measure operation and determine savings from VSDs over a five year period.
- The compressed air measure group was also analyzed for PG&E electric and SCE electric domains. This group has a low GRR for both PG&E and SCE. This is an area for particular attention in multiple areas, such as operating conditions and calculation methods. Closer attention should be given to more accurate calculation methodologies using AirMaster, and validating inputs and post-retrofit operating conditions (airflow, pressures and hours of use) to the greatest extent possible. Production should be used to normalize electrical use and savings.
- Boilers and heat recovery measures perform in line with the statewide gas average GRR for the PG&E gas and the SCG/SDG&E gas domains. Operating conditions such as hours of use, production rates, efficiencies and input temperatures were a frequent cause of discrepancies.
- For the HVAC controls group, the PG&E gas domain has a very high GRR (2.33), while the PG&E electric, SDG&E electric and SCG/SDG&E gas domains have among the lowest average GRRs of all groups analyzed. The PG&E electric and gas domains show a very wide spread, and calculation methods may be substantially improved as one way to rectify this divergence. *Data collection should center on the most important variables on which savings were based (e.g., temperatures, hours of use, etc.) post-retrofit and after systems have stabilized into their new mode of operation, and clearly identify when strategies have been discontinued.* In the SCG and SDG&E gas domain, baseline is a key reason for discrepancy. The SDG&E electric domain shows consistently low project level GRRs for this measure group.
- SDG&E electric also has low GRRs in the HVAC and whole building groups. These are areas for improvement. *SDG&E should make certain that opening conditions are as accurately represented as possible, that savings are adjusted in project documentation and tracking systems, that appropriate models are used, and that tracking systems contain the correct savings values in the whole building group. In addition, SDG&E should carefully review electric projects for eligibility, adherence to program rules and proper baseline determination.*
- The results for HVAC measures vary widely. While SDG&E electric showed low results overall, SCE project appeared to be performing well, until removing an outlier. PG&E

gas domain has a limited number of projects in the HVAC group. PGE gas appeared to be performing poorly, until removing an extreme point, upon which, PG&E electric performed at an average level. The effect of one or two projects needs to be carefully considered when considering program or measure-level performance. Results are intended to be directional and highlight areas for further investigation.

- For SDG&E electric, the refrigeration group is a strong performer with an average realization rate of 1.46. Several projects have a realization rate over 1.00. One likely cause is the use of conservative recommendations by SDG&E for complex refrigeration projects.
- Changes in operating conditions was the most frequent discrepancy factor noted across all measure groups, with the exception of the EC motors group.
- Changes in calculation methods was the second most common discrepancy factor and was noted for nearly all measure groups; this factor was particularly significant for the EC motors group, where the ex-post calculation method was substantially changed to account for ex-ante deficiencies. Different energy models or different methods within the same models were used, and different approaches, such as billing analysis as opposed to spreadsheet calculations were employed. Calculation method changes also included different spreadsheet calculations, but sometimes only minor changes were applied to the spreadsheets used for ex-ante calculations.
- Baseline issues appear to be concentrated in the HVAC, HVAC controls, and VSD groups. A relatively small number of projects where baseline emerged as a discrepancy factor were found in the pump rehabilitation and boiler groups. Baseline issues generally result in a project GRR less than one; however, of the 90 cases in which baseline was determined to be a primary or secondary discrepancy factor, 14 of these resulted in a greater ex-post estimate of savings.
- Many projects were incorrectly classified as early replacement projects, and ex-ante savings were found to be improperly based on a baseline assumed to be the in situ equipment. Projects with ER claims were commonly reclassified under the ex-post evaluation as normal replacement projects. Note that there were also cases in which natural replacement was claimed and used for ex-ante savings, but the evaluation determined that early retirement was appropriate and savings were estimated accordingly.
- Ineligible measures occur infrequently as a discrepancy factor. Projects with this issue can be found primarily in the HVAC, HVAC controls, compressed air, and boiler groups. Ineligible equipment was installed in 30 sampled projects. Some reasons why projects were found to be ineligible includes the following: program rules that do not allow repairs, like for like replacements, retrofit measures that did not exceed standard practice, other program rule violations, and operational changes (such as HVAC control measures involving temperature changes).

5.5 New Construction / Whole Building Programs and Projects

New construction is an area of special interest. The long term impact of decisions made regarding energy efficiency during new construction makes this a growing area of opportunity. The incremental costs of improvements made during construction are far less than the costs of retrofitting measures after the building is completed. Integrated design can further reduce costs by trading costs for load avoidance measures (efficient lighting, high efficiency glazing, etc.) for reduced HVAC system size and cost. The advanced pathway to higher building codes and standards can be a driving force for reducing the energy needs of new buildings. Consequently, this segment continues to be an area of high interest to CPUC and IOUs and complements the growing interest in the concept of zero net energy buildings.

This subsection describes the gross impact results for 43 new construction projects that were selected for M&V. The total ex-ante savings claimed for the 43 NRNC sites is 57,457,606 kWh, 8,213.6 kW and 1,434,128 therms whereas the total evaluated (ex-post) savings for these 44 NRNC sites is 46,645,505 kWh, 5,936 kW and 728,222 therms. The un-weighted gross realization rate for NRNC sites is 0.81 for the kWh savings, 0.72 for the kW savings and 0.51 for therm savings.

A subset of these projects were incented under the whole building approach (WBA) and used the Energy Pro Title 24 compliance software to claim the savings in the ex-ante estimate. The total ex-ante savings claimed for the 25 WBA sites is 14,487,964 kWh, 3,092.9 kW and 529,779 therms; the total evaluated (ex-post) savings for these 25 project sites is ,898,955 kWh, 1,911.2 kW and 295,617therms. The un-weighted gross realization rate for WBA sites is 0.61 for kWh, 0.61 for kW and 0.56 for therms.

While the results for the new construction electric claims are relatively high overall, for the WBA projects, electric realization rates are somewhat lower than the statewide average. Natural gas GRRs are significantly less than the statewide average for both new construction and the WBA subset of projects. These projects and a thorough analysis are presented in Appendix F; specific program recommendations resulting from the new construction gross impact findings are included in Chapter 8.

There were issues observed during the evaluation of the new construction measures (typically whole building projects). Some of the major issues identified for the whole building projects are discussed below.

- For all whole building sites with energy savings simulated with Energy Pro, the ex-ante annual energy savings were determined based on standard T24 schedules instead of the building's as-designed schedules.

- The system configuration modeled in the building simulation does not always match the actual as-built system configuration. The system configuration in the submitted ex-ante model is typically different from what the evaluator finds during the site visit. For example, the ex-ante proposed model of one site indicated that the building is conditioned by an air-cooled chiller; the site visit determined that this building is actually conditioned by DX split units.
- For a number of NRNC projects, the ex-ante calculations were based on optimal operating conditions that did not match the observed building operating conditions. During site visits, evaluators have observed buildings that are operating equipment using inefficient schedules, inappropriate control sequencing, and by-pass of automatic equipment controls.

5.6 MBCx Monitoring Based Commissioning Programs

This subsection describes IOU-sponsored projects implemented under Monitoring Based Commissioning (MBCx) programs at UC/CSU campuses and elsewhere.

The retro-commissioning and re-commissioning processes typically involve a diverse mix of measures, with a concentration in the HVAC end uses. The measures may entail physical equipment changes but often entail operational changes prohibited by other programs. Measure life is typically six years to reflect the operational nature of the change and the probable need for future retro-commissioning / operational optimization efforts. The conventional MBCx ex-ante impact estimation approach is limited to whole building energy analysis, with metering/monitoring at the building level, and no separate analyses are utilized to estimate the energy savings for individual measures.

Site specific savings for the sampled MBCx projects are presented in Table G-1. This table shows the ex-ante savings, ex-post savings, and gross realization rates (GRR) for kWh, kW and therms. The total ex-ante savings claimed for the 10 MBCx sites were 9,087,120 kWh, 1,068 kW and 772,466 therms whereas the total ex-post savings for these 10 sites were 5,305,117 kWh, 1,150 kW and 269,177 therms. The un-weighted gross realization rate for the MBCx sites was 58 percent for the kWh savings, 108 percent for the kW savings and 35 percent for the therms savings. The overall higher GRR for with the peak demand reduction is due to the fact that for many projects, the Ex-ante savings analysis did not estimate the peak demand reduction.

MBCx and other retro-commissioning programs are associated with some of the lowest GRRs in this evaluation. Full details of MBCx projects are contained in Appendix G.

A summary of the findings on with the present approaches for implementing the MBCx projects and estimating the energy savings include the following:

- Some buildings selected for MBCx projects do not meet the program's minimum Energy Use Intensity (EUI) requirements and receive RCx services in buildings with inadequate savings opportunities.
- Three months of pre and post metered data for establishing the baseline and post-case is often inadequate because it fails to cover the seasonal variations.
- Non-aligned data collection intervals (such as some operational parameters collected at different time intervals) and inappropriate data collection periods (such as collecting data during winter months for cooling loads) are sometimes utilized, distorting the energy model utilized for savings calculations.
- Inadequate details are provided on extraneous building load and system characteristics.
- Operation constraints of existing equipment that limit the ability to implement revised control sequences.
- Existing controls system may not be compatible with the control changes proposed.
- Reliability of the installed meters is sometimes questionable and the new meters installed did not meet the instrumentation accuracy criteria specified in the MBCx project guidelines.
- Statistically invalid regression models with poor R-Squared values that do not confirm to IPMVP Option C metric guidelines are sometimes used for savings estimation. The IPMVP considers a model valid if the R-Square value is more than 0.75.
- Pre- and post-MBCx functional tests required under the program rules are not carried out to document the system operational performance and to verify measure implementation.
- Energy models developed from short-term metered data are not calibrated against pre- and post-MBCx annual energy usage profiles.
- Measures with short life spans and limited persistence are often included in projects.

MBCx and other retro-commissioning programs are associated with some of the lowest GRRs in this evaluation. Full details of MBCx projects are contained in Appendix G.

6

NTG Results

These findings are excerpted from a separate Net-to-Gross (NTG) report, which provides a more in-depth reporting of NTG methods, data sources, findings and recommendations. More detailed findings are provided in Appendix D.

The methodology used to develop the individual, site-specific net-to-gross estimates is summarized in the Evaluation Plan provided previously.[1] Here, we present the weighted results both for each sampling domain and for selected programs where the findings are sufficiently robust.[2]

6.1 Number of Completed Surveys

A substantial number of NTG surveys were completed; 1,388 in total. Across all four IOUs, the number of completed surveys was roughly proportional to the population of completed projects for each domain.[3]

Table 6-1 below reports the number of sampled projects by utility represented by all of the surveys completed for IOU Core and Third Party Programs.

[1] http://www.energydataweb.com/cpucFiles/pdaDocs/814/WO33%20Research%20Plan%20Final%2012%2029.pdf

[2] Defined as having a minimum of eight completed surveys, representing all relevant strata.

[3] See the separate Net-to-Gross report for a series of sample design tables to validate this assertion.

Table 6-1: Completed Surveys for IOU Core and Third Party Programs

Utility/Fuel Sampling Domain	Number of Completed Surveys (N)		
	All Periods	BD+AD1	AD2+AD3
PG&E Electric	558	381	177
PG&E Gas	230	170	60
SCE Electric	367	227	140
SDG&E Electric	125	78	47
SDG&E/SCG Gas	108	75	33
Total	**1,388**	**931**	**457**

Included in this total are over 250 surveys completed for projects funded by Statewide Government/Institutional[4] and Local Government programs as shown in Table 6-2 below.

Table 6-2: Completed Surveys for Statewide and Local Government Programs

Utility/Fuel Sampling Domain	Number of Completed Surveys (N)
PG&E Statewide Govt/Institutional	72
PG&E Local Government	69
SCE Statewide Govt/Institutional	56
SCE Local Government	57
Total	**254**

Appendix D provides a series of detailed tables summarizing statistics for the population and net-to-gross sample completes used to develop the final weighted results for each sampling domain. Note that the net-to-gross sample is larger than the gross sample; in addition to gross sampled sites, it also includes a number of 'net-only' sites. For all sampling domains, a large number of surveys were completed, representing very high percentages of the total population and providing for robust results across all sample strata.

6.2 Weighted NTG Results

The methodology used to develop the individual, site-specific net-to-gross estimates is summarized in Section 4. Here, we present the weighted results for each sampling domain. To produce an estimate of the net-to-gross ratio (NTGR), the individual NTGRs for each of the applications in the sample were weighted by the size of the ex-ante impacts associated with the application and the proportion of the total sampling domain impacts represented by each sampling stratum.

4 Comprised of the following programs: UC/CSU, CCC, DOC

The tables below present statistics for the population and net-to-gross sample completes used to develop the final weighted results for each sampling domain. For all utility/fuel domains, a large number of surveys were completed, providing for robust results across most programs and program groupings.

Weighted NTGRs were calculated for each size stratum within each program, enabling closer examination of the factors driving program level NTGRs. In general, the large number of completed surveys enables reporting for a sizable number of programs and program groupings. In some cases, the number of completed surveys within a stratum was either zero or too small to support a weighted estimate, and such cases are noted.

Note that the Final NTGR values in tables 6-3, 6-7, 6-9 and 6-11 below are based on the removal of 9 projects for the limited purpose of calculating an NTGR Adjustment Factor. As described in Chapter 4, this was due to the potential overlap for these sites between the NTG for the project and the Gross ISP or Dual Baseline determinations in the gross impact analysis. For all IOU-fuel domains except PG&E Gas,[5] the improvement in IOU-fuel domain level NTGRs from these removals was very slight, on the order of one to two percent. Due to the small number of project NTGs removed, the very small overall adjustment, and the smaller sample sizes at the program level, only the primary results for the IOU-fuel domain were adjusted for this overlap.

- General NTGR Observations
  - In general, the weighted NTGRs for the **electric fuel domain have not improved compared to evaluated values** from the PY2006-2008 evaluation of industrial programs for both PG&E and Southern California Edison. At the level of IOU-fuel sampling domain, the final NTGRs range from 0.46 to 0.50 over the complete cycle. Nevertheless, certain niche programs experience much lower NTGRs.
  - However, there has been a **significant improvement in the weighted NTGRs for PG&E industrial gas programs** over those seen in the PY2006-2008 industrial evaluation.[6] In that evaluation, the gas NTGR is particularly low (0.31), but has more than doubled for both the PG&E Core Calculated – Industrial and the All PG&E Core programs groupings. However, certain programs such as PGE2225 (Nexant's Refinery Energy Efficiency program), PGE21042 (New Construction) and the Local Government Partnership (the PG&E Energy Watch program group) continue to experience high free ridership. In addition, note that the NTGR for PG&E agriculture and food processing programs in this evaluation is similar to that in the PY2006-2008 agriculture and food processing evaluation[7] (0.63).

[5] PG&E Gas had zero projects removed for the NTG Adjustment Factor and therefore the results were unaffected.

[6] http://www.calmac.org/publications/PG&E_Fab_06-08_Eval_Final_Report.pdf

[7] http://www.calmac.org/publications/PG%26E_Ag-Food_Eval_Report_V1_021010ES.pdf

Below is a more detailed summary of NTGR results by utility and fuel domains.

6.2.1 PG&E Electric

In general, NTGRs for electric projects developed through PG&E programs are in line with those reported in the PY2006-2008 evaluations (kWh NTGR = 0.45, kW NTGR = 0.44). The weighted NTGR across all programs and projects was 0.47, and there was little movement seen between the BD+AD1 period and the AD2+AD3 period. Table 6-3 below reports NTGRs across all of the program groups represented by the completed surveys.

Table 6-3: Weighted Net-to-Gross Ratios for PG&E – Electric

Program Sampling Strata	All Periods	BD+AD1	AD2+AD3
1	0.55	0.51	0.66
2	0.46	0.47	0.42
3	0.46	0.48	0.42
4	0.44	0.43	0.46
5	0.44	0.42	0.48
Weighted NTGR	**0.46**	**0.46**	**0.47**
90 Percent CI	0.446 to 0.483	0.44 to 0.485	0.438 to 0.497
Relative Precision	0.04	0.05	0.06
n NTGR Completes	558	381	177
N Sampling Units	6,994	4,706	2,288
ER	0.59	0.60	0.53
NTGR Adjustment Factor	1.01		
Final NTGR	**0.47**		

PG&E program domain-specific NTGR values varied widely, as shown in Table 6-4 below.

Table 6-4: Weighted Net-to-Gross Ratios for PG&E – Electric Programs and Program Groupings

Program/Program Group Result	Number of Completed Surveys	Net-to-Gross Ratio
PG&E - Electric		
Core Programs		
Calculated Incentives - Comm Ind Ag	195	0.47
Calculated Incentives - Commercial	86	0.57
Calculated Incentives - Industrial	44	0.28
Calculated Incentives -Agricultural	65	0.47
Non-Core Programs		
PGE21042: New Construction	19	0.46
PGE20135: Pump Efficiency Services	36	0.36
PGE2222: Energy Eff. Oil and Gas	46	0.37
PGE2223: Heavy Industry Energy Efficiency	27	0.62
PGE2225: Refinery Energy Efficiency	9	0.66
PGE2228/2203: RCx Group	12	0.62

Within the Core Calculated program group, values ranged from a low of 0.28 (Core Calculated Incentives Industrial) to a high of 0.57 (Core Calculated Incentives Commercial). Within the Third Party program group, values varied from a low of 0.37 (PGE2222 Energy Efficiency for Oil and Gas) to a high of 0.66 (PGE2225 Refinery Energy Efficiency Program). NTGRs for both the Core Calculated Incentives Industrial program and the third party PGE2225 Refinery Energy Efficiency Program were much lower due to the presence of legacy pump-off controller projects that originated in PY2006-2008 and have become standard practice now.

The very highest performing programs in this domain - those with among the highest NTGRs - still had NTGR values at or below 0.66. These programs included: Calculated Incentives Commercial (NTGR = 0.57), Heavy Industry Energy Efficiency (NTGR = 0.62), Refinery Energy Efficiency (NTGR = 0.66) and Retrocommissioning (NTGR = 0.62).

The very lowest performing programs - those with among the lowest NTGRs – included Calculated Incentives Industrial (NTGR = 0.28), Agricultural Pump Efficiency (NTGR = 0.36), and Energy Efficiency for Oil and Gas (NTGR = 0.36). Several other programs have NTGRs below 0.50.

6.2.2 PG&E Gas

In contrast to its electric projects within its programs, NTGRs for PG&E custom gas projects within its programs are significantly improved from the PY2006-2008 industrial evaluation[8]

[8] Ibid.

where the NTGR for gas projects averaged 0.31.[9] As an example, the NTGR for gas projects across all PG&E Core programs (0.56) is nearly twice as high as that in the PY2006-2008 industrial evaluation. Table 6-5 below reports calculated NTGRs across all of the program groups represented by the completed surveys. In addition, NTGRs for the largest stratum 1 and 2 projects are very low, and similar to levels found in the PY2006-2008 industrial evaluation. However, NTGRs for the medium and small projects that populate strata 3, 4 and 5 are significantly improved from the levels found in the PY2006-2008 industrial evaluations.

Table 6-5: Weighted Net-to-Gross Ratios for PG&E – Gas

<table>
<tr><th>Program Sampling Strata</th><th>All Periods</th><th>BD+AD1</th><th>AD2+AD3</th></tr>
<tr><td>1</td><td rowspan="2">0.33</td><td rowspan="2">0.33</td><td>N/A*</td></tr>
<tr><td>2</td><td>N/A*</td></tr>
<tr><td>3</td><td>0.71</td><td>0.71</td><td rowspan="2">0.48</td></tr>
<tr><td>4</td><td>0.57</td><td>0.60</td></tr>
<tr><td>5</td><td>0.56</td><td>0.55</td><td>0.60</td></tr>
<tr><td>Weighted NTGR</td><td>0.56</td><td>0.54</td><td>0.59</td></tr>
<tr><td>90 Percent CI</td><td>0.532 to 0.582</td><td>0.514 to 0.567</td><td>0.533 to 0.647</td></tr>
<tr><td>Relative Precision</td><td>0.05</td><td>0.05</td><td>0.10</td></tr>
<tr><td>n NTGR Completes</td><td>230</td><td>170</td><td>60</td></tr>
<tr><td>N Sampling Units</td><td>1,270</td><td>859</td><td>411</td></tr>
<tr><td>ER</td><td>0.46</td><td>0.44</td><td>0.49</td></tr>
</table>

* No projects

Program-specific NTGR values range from 0.35 to 0.63 as shown below in Table 6-6.

9 Note that the PG&E custom gas projects in this evaluation include a commercial component, in contrast to the PY2006-2008 industrial evaluation.

Table 6-6: Weighted Net-to-Gross Ratios for PG&E – Gas Programs and Program Groupings

Program/Program Group Result	Number of Completed Surveys	Net-to-Gross Ratio
PG&E - Gas		
Core Programs		
Calculated Incentives - Comm Ind Ag	97	0.63
Calculated Incentives - Commercial	40	0.55
Calculated Incentives - Industrial	27	0.63
Calculated Incentives -Agricultural	30	0.61
Non-Core Programs		
PGE21042: New Construction	7	0.39
PGE2223: Heavy Industry Energy Efficiency	12	0.57
PGE2225: Refinery Energy Efficiency	5	0.35
PGE2228/2203: RCx Group	8	0.63

The lowest values are for the Local Government Partnership and PG&E Energy Watch programs (NTGRs = 0.20), the Refinery Energy Efficiency Program (NTGR = 0.35) and the Savings by Design new construction program (NTGR = 0.39); however, note the sample sizes for all three of these programs are fairly small. Among the strongest performing programs, those with the highest NTGRs, are: Other Third Party Programs[10] (NTGR = 0.63), Core Industrial (NTGR = 0.63), and Retrocommissioning (NTGR = 0.63). The Core Commercial-Industrial-Agricultural grouping also has an NTGR of 0.63, approaching the default ex ante NTGR of 0.64.

6.2.3 SCE Electric

NTG ratios for SCE's programs (all resulting in electric savings) are somewhat lower than those for SCE's industrial programs in PY2006 – 2008, which had an NTGR of 0.63. In the current evaluation, the weighted NTGRs by stratum and across all size projects are clustered around 0.50 as shown in Table 6-7 below.

[10] Includes PGE2186 Enhanced Automation Initiative, PGE2132 C/I Boiler Efficiency , PGE2234 Comprehensive Food Process Audit & Resource Efficiency , PGE2193 School Energy Efficiency, PGE2190 Lodging Savers, PGE2209 Ozone Laundry Energy Efficiency, and PGE2205 Casino Green.

Table 6-7: Weighted Net-to-Gross Ratios for SCE – Electric

Program Sampling Strata	All Periods	BD+AD1	AD2+AD3
1	0.48	0.46	0.49
2	0.52	0.52	0.47
3	0.52	0.52	0.53
4	0.47	0.45	0.52
5	0.45	0.45	0.45
Weighted NTGR	**0.49**	**0.48**	**0.50**
90 Percent CI	0.47 to 0.504	0.461 to 0.503	0.465 to 0.527
Relative Precision	0.04	0.04	0.06
n NTGR Completes	367	227	140
N Sampling Units	3,052	1,737	1,315
ER	0.44	0.43	0.47
NTGR Adjustment Factor	1.02		
Final NTGR	**0.50**		

By program category, NTGR results are similar, ranging from a low of 0.40 for core commercial to a high of 0.51 for the Other Third Party category[11], as indicated in Table 6-8 below.

Table 6-8: Weighted Net-to-Gross Ratios for SCE – Electric Programs and Program Groupings

Program/Program Group Result	Number of Completed Surveys	Net-to-Gross Ratio
SCE - Electric		
Core Programs		
Calculated Incentives - Comm Ind Ag	162	0.47
Calculated Incentives - Commercial	59	0.40
Calculated Incentives - Industrial	39	0.54
Calculated Incentives -Agricultural	64	0.44
Non-Core Programs		
SCE-SW-005A - New Construction	38	0.45
Other Third Party	54	0.51

The Other Third Party category is characterized by moderate program influence, with an NTGR of 0.51 across the 54 projects evaluated. Within the core program category, Industrial is the

[11] The Other Third Party category consisted of 54 projects from the following programs: Nonmetallic Minerals and Products (SCE-TP-016), Oil Production (SCE-TP-020), Food & Kindred Products (SCE-TP-013), Primary and Fabricated Metals (SCE-TP-014), Retail Energy Action Program (SCE-TP-025), Comprehensive Petroleum Refining (SCE-TP-019), Healthcare EE Program (SCE-TP-006), Management Affiliates Program (SCE-TP-031), and Commercial Utility Building Efficiency (SCE-TP-026).

strongest performer with an NTGR of 0.54, while Commercial with an NTGR of 0.40 was the weakest.

6.2.4 SDG&E Electric

NTGRs for SDG&E's electric projects are similar in magnitude to SCE's, averaging 0.46 across all periods studied. Results by period show some variation. Table 6-9 below reports verified NTGRs across all of the program groups represented by the completed surveys.

Table 6-9: Weighted Net-to-Gross Ratios for SDGE – Electric

Program Sampling Strata	All Periods	BD+AD1	AD2+AD3
1	0.41	0.48	0.29
2	0.49	0.48	0.51
3	0.43	0.38	0.55
4	0.48	0.47	0.48
5	0.47	0.43	0.56
Weighted NTGR	**0.45**	**0.45**	**0.47**
90 Percent CI	0.427 to 0.482	0.415 to 0.486	0.437 to 0.505
Relative Precision	0.06	0.08	0.07
n NTGR Completes	125	78	47
N Sampling Units	1,469	615	854
ER	0.43	0.45	0.31
NTGR Adjustment Factor	1.02		
Final NTGR	**0.46**		

Only three programs or program groups had sufficient sample to be able to report on: SDGE core, SCGE new construction and SDGE3117 (Nonresidential BID). These results are shown below in Table 6-10.

Table 6-10: Weighted Net-to-Gross Ratios for SDGE – Electric Programs and Program Groupings

Program/Program Group Result	Number of Completed Surveys	Net-to-Gross Ratio
SDG&E - Electric		
Core and Non-Core Programs		
Calculated Incentives - Comm Ind Ag	56	0.45
SDGE3118 - New Construction	14	0.33
SDGE3117 - Local Nonresidential BID program	55	0.49

The Nonresidential BID program had the strongest performance of the three (NTGR = 0.49) while the new construction program performed weakest (NTGR = 0.33). However, note that there are a small number of completed interviews for the New Construction program. The Core Calculated Incentives programs NTGR of 0.45 demonstrated a medium-low program influence level.

6.2.5 SDG&E and SCG Gas

NTGRs for SDG&E and SCG gas programs show somewhat more variation across the different program grouping categories. The weighted NTGR across all programs and projects is 0.50, with some improvement noted for the AD2+AD3 period over the BD+AD1 period, as shown in Table 6-11 below.

Table 6-11: Weighted Net-to-Gross Ratios for SDGE/SCG – Gas

Program Sampling Strata	All Periods	BD+AD1	AD2+AD3
1	0.45	0.37	N/A (No projects)
2			0.71
3	0.51	0.52	0.50
4	0.56	0.53	0.62
5	0.51	0.51	0.52
Weighted NTGR	**0.50**	**0.45**	**0.60**
90 Percent CI	0.423 to 0.571	0.368 to 0.523	0.567 to 0.632
Relative Precision	0.15	0.17	0.05
n NTGR Completes	108	75	33
N Sampling Units	1,077	444	633
ER	0.99	1.00	0.19
NTGR Adjustment Factor	1.00		
Final NTGR	**0.50**		

By program category, SDG&E's Nonresidential BID program (NTGR = 0.67) had the strongest performance as shown below in Table 6-12.

Table 6-12: Weighted Net-to-Gross Ratios for SDGE and SCG – Gas Programs and Program Groupings

Program/Program Group Result	Number of Completed Surveys	Net-to-Gross Ratio
SDG&E - Gas		
Non-Core Programs		
SDGE3117 - Local Nonresidential BID program	22	0.67
SCG - Gas		
Core Programs		
Calculated	62	0.48
Deemed	10	0.55

Among SCG's programs, its Deemed (NTGR = 0.55) and Core (NTGR = 0.48) programs exhibited moderate program influence.

6.2.6 Statewide Government and Institutional Partnerships - Electric

NTGRs for statewide government and institutional partnership programs are similar in magnitude to those for IOU-based programs. Across all evaluated programs and projects, an average NTGR of 0.56 was achieved. By IOU, evaluated NTGRs are very similar, as shown below in Table 6-13.

Table 6-13: Weighted Net-to-Gross Ratios for Statewide Government and Institutional Partnership programs - Electric

Program Results	Statewide Govt	PGE Govt & Institutional	SCE Govt & Institutional
Weighted NTGR	**0.56**	**0.55**	**0.57**
90 Percent CI	0.536 to 0.584	0.518 to 0.592	0.54 to 0.594
Relative Precision	0.04	0.07	0.05
n NTGR Completes	128	72	56
N Sampling Units	434	308	126
ER	0.35	0.39	0.29

By program category, results are more divergent. The statewide community college partnership program has a somewhat lower NTGR (0.47), while the UC/CSU NTGR is moderately higher (0.56). The local government program category has the least favorable result, with an NTGR averaging 0.45 across the state.

6.3 Key Factors Influencing NTGRs

Behind each of the NTGRs calculated for each project is a host of contextual factors that may have influenced the project, either directly or indirectly. The key contextual factors were first examined within each project, and then summarized across all evaluated projects within a given program or program grouping. The intent was to look more deeply, beyond the numerical responses used in the NTGR algorithm, into the qualitative factors that influenced the project decision making. The tables below provide the results of this analysis across the programs and program groupings for which the weighted NTG factors were developed. The following are general themes and observations across these analyses:

- **Corporate policy is a major driver of most projects.** Corporate standard practice is nearly-universal as a motivating factor. Related to this is the strong presence of corporate environmental policies.
- For better-performing programs and program groups (those with the majority of NTGRs in the Medium High category) none of the other, non-Corporate policy factors examined[12] were present. This suggests that these programs are more effective at steering clear of many of the issue areas that plague low program influence.
- For weaker programs and program groups with the majority of NTGRs in the low and medium low categories, **there are one or more other strong drivers present that may be contributing to poor performance.** For example:
  - For PGE2222, Energy Efficiency Services for Oil Production, a majority of projects have **automation benefits in addition to energy savings**. Also, a significant percentage of projects were already in an **advanced stage,** i.e., program implementers arrived late in the decision making process and offered incentives for projects that had already been decided upon**.**
  - New construction projects have many non-energy efficiency drivers. For PGE21042, new construction, a significant percentage of projects are being implemented by **firms advanced in their adoptions of energy efficiency**, including national chains and big box stores.
  - **Replacement of failing equipment** is a key motivator for a large number of projects in the SCE Core, Calculated, and Other Third Party (3P) program groups.
  - **Environmental compliance** has a strong presence in the decisions of a significant number of projects for the SCE UC/CSU institutional partnerships, SDGE3117 Nonresidential BID, and SDGE3118 new construction programs.

[12] Such as environmental compliance, strong non-energy benefits, projects motivated by facility expansion or modernization, among others.

Table 6-14: NTG Reasons by Project: PG&E Electric and Gas

	PGE Core Comm Ind Ag	Energy Efficiency Services for Oil Production	New Construction	Heavy Industry	RCx Group	Other 3P PGE
		PGE2222	PGE21042	PGE2223	RCx Group	
Distribution of NTGRs						
High - 0.76 to 1.00	9%	0%	10%	5%	14%	13%
Medium High- 0.51 to 0.75	30%	7%	24%	55%	50%	42%
Medium Low- 0.26 to 0.50	49%	50%	43%	32%	36%	38%
Low - 0.00 to 0.25	12%	43%	24%	8%	0%	7%
Program/Program Grouping NTGR - Electric	**0.47**	**0.37**	**0.46**	**0.62**	**0.62**	**0.47**
Program/Program Grouping NTGR - Gas	**0.63**	**N/A**	**0.39**	**0.57**	**0.63**	**0.68**
Key Project Drivers						
Project Maturity						
Project is in the capital and/or operating budget	5%	33%	11%	3%	7%	0%
Equipment has already been ordered	1%	22%	6%	0%	0%	0%
Corporate Policy/Practice						
Measure is part of corporate standard practice	67%	46%	61%	68%	86%	62%
Measure is installed elsewhere in company, in places that do not offer rebates	14%	41%	22%	3%	7%	1%
Company has environmental policy in place	53%	22%	78%	49%	71%	52%
Energy Efficiency A Secondary, not Primary, Benefit						
Measure automates existing manual processes	11%	65%	11%	14%	7%	13%
Measure improves workplace quality	14%	0%	33%	0%	7%	14%
Environmental Compliance						
Measure is associated with environmental compliance (e.g., pollution reduction)	6%	0%	0%	3%	0%	7%
Market Segment						
Measure is installed by a market segment that is ahead of curve on energy efficiency	10%	33%	17%	0%	0%	4%
Measure is installed by national chain/big box firm	10%	0%	22%	0%	0%	6%
Project Cost vs. Rebate						
Rebate is very small % of overall project cost	7%	28%	11%	16%	0%	1%
Project Context						
Measure is part of an expansion/remodeling	16%	26%	28%	14%	7%	7%
Measure installed to replace failing equipment	20%	4%	0%	5%	29%	18%

Table 6-15: NTG Reasons by Project: SCE Electric

	SCE Core Comm Ind Ag	Calculated Incentives Commercial	Calculated Incentives Industrial	Calculated Incentives Agricultural	Core New Construction	Institutional Partnerships - UC/CSU	Other 3P SCE Group
		SCE-SW-002B	SCE-SW-003B	SCE-SW-004B	SCE-SW-005A	SW UC/CSU	
Distribution of NTGRs							
High - 0.76 to 1.00	6%	2%	13%	5%	3%	0%	6%
Medium High- 0.51 to 0.75	36%	28%	51%	34%	29%	59%	37%
Medium Low- 0.26 to 0.50	49%	57%	31%	53%	63%	38%	48%
Low - 0.00 to 0.25	9%	14%	5%	8%	5%	3%	9%
Program/Program Grouping NTGR	**0.47**	**0.40**	**0.54**	**0.44**	**0.45**	**0.57**	**0.51**
Key Project Drivers							
Project Maturity							
Project is in the capital and/or operating budget	1%	2%	0%	2%	8%	0%	0%
Equipment has already been ordered	0%	0%	0%	0%	5%	0%	0%
Corporate Policy/Practice							
Measure is part of corporate standard practice	80%	83%	67%	84%	70%	56%	76%
Measure is installed elsewhere in company, in places that do not offer rebates	3%	5%	3%	2%	5%	0%	0%
Company has environmental policy in place	54%	57%	59%	48%	73%	81%	63%
Energy Efficiency A Secondary, not Primary, Benefit							
Measure automates existing manual processes	9%	17%	3%	5%	0%	0%	11%
Measure improves workplace quality	7%	12%	13%	0%	11%	34%	9%
Environmental Compliance							
Measure is associated with environmental compliance (e.g., pollution reduction)	2%	7%	0%	0%	14%	28%	2%
Market Segment							
Measure is installed by a market segment that is ahead of curve on energy efficiency	4%	16%	0%	0%	0%	0%	2%
Measure is installed by national chain/big box firm	6%	5%	8%	0%	0%	0%	6%
Project Cost vs. Rebate							
Rebate is very small % of overall project cost	3%	7%	3%	0%	11%	0%	0%
Project Context							
Measure is part of an expansion/remodeling	7%	10%	10%	2%	19%	13%	6%
Measure installed to replace failing equipment	23%	21%	15%	30%	0%	3%	39%

Table 6-16: NTG Reasons by Project: SDG&E/SCG Electric and Gas

	SCG Core Calculated	SCG Deemed	SCG Third Party	SDGE Core Calculated	SDGE Local	SDGE New Construction
	SCG3602/07/11	**SCG3612**	**SCG3663**	**SDGE3100/05/09**	**SDGE3117**	**SDGE3118**
Distribution of NTGRs						
High - 0.76 to 1.00	8%	0%	50%	2%	4%	0%
Medium High- 0.51 to 0.75	44%	70%	50%	48%	43%	20%
Medium Low- 0.26 to 0.50	42%	30%	0%	37%	38%	73%
Low - 0.00 to 0.25	6%	0%	0%	13%	16%	7%
Program/Program Grouping NTGR - Electric	**N/A**	**N/A**	**N/A**	**0.45**	**0.49**	**0.33**
Program/Program Grouping NTGR - Gas	**0.49**	**0.55**		**N/A**	**N/A**	**N/A**
Key Project Drivers						
Project Maturity						
Project is in the capital and/or operating budget	2%	0%	0%	0%	9%	7%
Equipment has already been ordered	0%	0%	0%	0%	0%	0%
Corporate Policy/Practice						
Measure is part of corporate standard practice	55%	90%	100%	90%	59%	80%
Measure is installed elsewhere in company, in places that do not offer rebates	8%	20%	0%	2%	9%	0%
Company has environmental policy in place	50%	40%	50%	70%	70%	67%
Energy Efficiency A Secondary, not Primary, Benefit						
Measure automates existing manual processes	10%	0%	0%	13%	13%	0%
Measure improves workplace quality	10%	0%	0%	10%	14%	20%
Environmental Compliance						
Measure is associated with environmental compliance (e.g., pollution reduction)	11%	10%	0%	0%	25%	27%
Market Segment						
Measure is installed by a market segment that is ahead of curve on energy efficiency	3%	0%	0%	7%	0%	7%
Measure is installed by national chain/big box firm	8%	0%	0%	0%	2%	0%
Project Cost vs. Rebate						
Rebate is very small % of overall project cost	10%	0%	0%	0%	0%	13%
Project Context						
Measure is part of an expansion/remodeling	15%	0%	0%	3%	16%	20%
Measure installed to replace failing equipment	0%	0%	0%	0%	0%	0%

7

Lower Rigor Assessment Results

7.1 Introduction

This chapter discusses the results of the Lower Rigor Assessment (LRA) effort. The goal of the LRA effort is to provide qualitative, cost-effective, program-specific, impact-oriented findings and feedback for a wider range of programs than would otherwise be possible with a traditional M&V approach. This was accomplished by supplementing the 200 BD period and 236 AD period projects selected for the more rigorous M&V-based gross impact evaluation with an additional 100 sites from the BD period that received only a lower rigor assessment. These points were targeted at programs of interest to ensure that program groups received a minimum combined M&V plus LRA only sample size for the LRA assessment. The LRA process entailed a "desk review" of project application paperwork received from the IOU. The lead evaluator for each project answered a standardized set of 17 questions in an electronic form to assess the quality of the documentation provided, the adherence to program rules and guidelines, and the appropriateness of savings estimation techniques. This section presents LRA results for all assessments and compares results across IOUs, specific IOU programs, IOU-specific program domains, and statewide program domains. More detailed results can be found in Appendix E including a description of the program groups, details of the LRA Form, and an explanation of the issue areas assessed.

LRA results are assessments of program performance that are qualitative in nature and are not necessarily indicative of impact evaluation findings and quantitative results. Only a weak correlation (e.g., R^2=0.04 for kWh) between the overall assessment of the quality of the project based on the LRA overall assessment score[1] and the project gross realization rate was found. While the positive relationship between the two methods is noted, a more in-depth review of the correlation was not pursued.

In instances where LRAs were conducted prior to full M&V, project-level LRA findings inform the M&V data collection process and prepare the on-site engineer to better understand the strengths and weaknesses of the project. The LRA is an independent analysis of the merits of the project and is distinct from the development of project-specific ex-post gross impact estimates.

1 Appendix E includes a detailed description of the LRA overall assessment score.

7.2 Lower Rigor Assessment Results

This report presents unweighted lower rigor assessment results by segment: IOU fuel sampling domain, program, and various program groupings. This section begins by briefly reviewing the assessment process, then discusses overall portfolio-level results, and finally provides a detailed segment review covering the specific qualitative findings identified in the lower rigor assessment process. These more detailed results compare like programs, program groups or other segment-level groupings, discussing program assessments relative to other portfolio results.

7.2.1 Overview of the LRA Performance Assessment

This section presents results and highlights statistically significant variants versus average performance of the complete "portfolio" sample. This serves to identify areas of success or areas of concern by segment, such as program and program grouping.

This comprehensive lower rigor assessment effort of custom impact projects is a first of its kind and therefore there are no results from prior evaluation cycles that provide a basis for comparison. This effort builds upon efforts previously discussed in the WO033 Interim Report[2] and the WO12 IOU Core Calculated Program Group Report[3] and attempts to establish a baseline for future evaluation cycles. Of note, the 2013-14 CPUC evaluations involve project assessments that can inform relative improvement or decline in program performance for each issue area examined.

LRA performance is generally assessed using "Quality of Implementation" metrics, as discussed in Appendix E, which were used to assess IOU documentation and provide qualitative project-by-project reporting on key issue areas. Assessment issue areas were scored as being either "good," "neutral," or "poor." However, the results reported here emphasize areas for improvement based on the percentage of a given segment receiving a "poor" assessment score for any given issue. While the intent is not to focus on the negative per se, a predominance of "poor" scores is a more definitive indicator of areas needing improvement than a deficiency of "good" or "neutral" scores. More specifically, "good" was not necessarily distinguishable from "neutral" in terms of being indicative of adherence to protocol. Instead, a "good" assessment indicated that the evaluating engineer noted the criteria to be a "significant" finding in that it was particularly well-documented or otherwise exemplary as compared to responses that the evaluator found with other projects. Thus, it was the relative quality and completeness of the information provided for a particular project as compared to the engineers' experiences with other projects that led to the selection of "good" rather than "neutral." A "neutral" score meant generally that there was no documentation which indicated that the issue "failed" the assessment,

---

2 http://www.energydataweb.com/cpucFiles/pdaDocs/901/WO033_Interim_Report_Final%2001%2016%2013.docx

3 http://www.energydataweb.com/cpucFiles/pdaDocs/963/Non_Res_Core_Calculated_Prog_Assess.pdf

and when in doubt the evaluator gave such projects the benefit of the doubt with a “neutral” rating instead of “poor.” Thus, “good” and “neutral” generally represent the percent of projects receiving a “pass” whereas “poor” indicates there is a significant documented shortcoming, issue, or problem.

In general the BD period results reflect LRA-based findings following an initial review of IOU-provided project documentation. Updates to the LRAs, if needed, were applied any time new IOU-supplied information was made available, for example, after receiving response(s) to any follow-up data requests submitted by the evaluation team. AD period results reflect LRA-based findings that are informed by full M&V activities (including site visits and analysis). To better understand how assessments may change during the various stages of the M&V process, the stage in the M&V process that each update to the LRA occurred in was recorded. This allowed for a comparison between the two versions of AD period LRA results, and this comparison is presented in Section 7.4.

The results presented in this section highlight the boundary that distinguishes a “typical” program from one that stands out either as “above average” or as “below average.” The boundary is set using the 90 percent confidence interval of the “percent poor” responses for the complete portfolio sample, denoted as All LRA Average. Programs or segment groupings showing a “percent poor” score for a given issue that falls below the lower bound of the 90 percent confidence interval (fewer poor assessments) are flagged as “above average” in that area. Programs with “percent poor” greater than the upper bound of the 90 percent confidence interval are those with performance “below average” in a given issue area (i.e., more poor assessments). Results are presented for specific issue areas where there are eight or more completed project assessments for a given issue area. Cells with zero “percent poor” indicate that there are no “poor” assessments for that area but this should **not** be used as an indication of significant good performance unless the cell is also shaded in light grey. For the common issue areas, fewer than 10 responses is indicated with “small sample” or “zero sample,” but for the low-frequency issue areas, the threshold is eight points.

7.2.2 Overall Portfolio Results

Table 7–1 below provides context for program and program group performance by showing the aggregate results across all sample points. For each of the 17 issue areas, the number of contributing assessments and the percent scoring “good,” “neutral,” and “poor” are shown. Sample sizes vary by issue area due to applicability of the issue and whether or not assessments were feasible for each project.

In order to better identify and highlight areas for improvement, two applications of the ”percent poor” results were derived, one for issue areas that are generally applicable to most projects and a second for low frequency issue areas. “Low frequency” issue areas are those areas with fewer than 100 responses. Both results are shown in Table 7–1 and differ with respect to the denominator that is applied in order to derive each percentage.

For the remaining issue areas with more than 100 responses, the calculation of "percent poor" is derived using the total sample size denominator (536). This percentage is also referred to as the overall denominator with the initials OD. This choice of denominator serves to better identify the frequency of occurrence in the sample, expressed as a percentage of all projects assessed.

For the low frequency issue areas the denominator applied is the sum of the valid responses to each issue area. Valid responses are good, neutral, and poor, and so this percentage type is referred to by the initials GNP. The percentages calculated in this way restrict the denominator value and thereby bring greater fidelity to the GNP results.[4] All result tables presented in this report use the GNP denominator for "percent poor" for the low frequency issues. "Percent good" and "percent neutral" results are also derived here using the GNP denominator but are only featured in Table 7–1.

The low frequency issue areas include: *appropriate HVAC interactive effects calculation method*, *appropriate non-HVAC interactive effects calculation method, multiple IOU fuel impacts properly accounted for (includes fuel switching and cogeneration), fuel switching supported with three prong test*, and *non-IOU fuel and ancillary impacts of project properly accounted for (cogen/waste heat recovery/refinery gas, etc.).* Please note that *italic* typeface is used in this chapter and the LRA Appendix when referring to specific *issue areas* to help distinguish between specific findings that relate to values in the tables and general discussion of those findings.

The magnitude of the GNP-based "percent poor" results by issue area coincides with prior 2006-2008 CPUC program evaluation findings and results in the custom program area. Table 7–1 shows poor results in *proper accounting of non-IOU fuel and ancillary project impacts*, *use of the three prong test*, *accounting for multiple IOU fuels*, and *treatment of early replacement claims and use of a valid RUL/EUL approach*. These continue to be areas where the IOUs need to improve considerably and where past evaluation findings and recommendations still have not been properly addressed by the IOUs to improve program performance.

[4] It should be noted that the use of the OD calculation applied to low frequency issue areas yields percentages that are quite low and do not convey the magnitude of poor results associated with the projects where a given consideration is relevant.

Table 7–1: Lower Rigor Assessments for Complete Portfolio Sample, Percent "Good," "Neutral," & "Poor"

Issue Assessed	Assessment Results (n)	Assessment Results (%)			
		Good	Neutral	Poor (OD)[5]	Poor (GNP)
Project Documentation and Specification					
IOU application documentation complete and accurate	513	39%	33%	26%	28%
IOU tracking data complete and accurate	520	38%	42%	20%	20%
Project utilized pre-installation M&V	390	43%	27%	22%	30%
Appropriate baseline	507	64%	14%	21%	22%
Early replacement claim: Valid RUL / EUL approach used	249	48%	12%	18%	40%
Appropriate Calculation Method					
Appropriate impact calculation method	492	43%	34%	21%	23%
All relevant inputs considered	458	63%	16%	18%	21%
Adequate values for all inputs	470	37%	40%	21%	23%
Appropriate HVAC interactive effects calculation method [6]	67	55%	12%	4%	33%
Appropriate non-HVAC interactive effects calculation method	77	62%	17%	3%	21%
Project utilized post-installation M&V	505	33%	34%	30%	32%
Compliance with Program Rules					
Measures are IOU program eligible	520	84%	13%	3%	3%
Measures exceed code or industry standard practice	473	75%	12%	11%	13%
Multiple IOU fuel impacts properly accounted for (includes fuel switching and cogeneration)	44	43%	5%	4%	52%
If applicable, fuel switching supported with three prong test	16	31%	0%	2%	69%
Non-IOU fuel and ancillary impacts of project properly accounted for (cogen/waste heat recovery/ refinery gas, etc.)	94	27%	0%	13%	73%
Customer installation meets all program rules	518	75%	12%	13%	13%

The OD-based "percent poor" results best emphasize the frequency of occurrence for a given issue area across all domains. Table 7–1 indicates that the *compliance with program rules* category is the area of least concern. The most frequently occurring issue area with roughly 30

[5] Generally, the total number of "poor" assessments is used for performance comparison.

[6] To provide greater fidelity in the performance score for low-frequency issue areas, the denominator for the five rows of the table marked with this footnote include only those LRAs where the issue was applicable to the project and received a "good," "neutral," or "poor" (GNP) assessment.

percent rated “poor” is appropriate use of *post-installation M&V* followed by lack of *completeness and accuracy of IOU documentation* at 26 percent.

7.2.3 Program Performance Summary

This section presents an overarching summary of the LRA results. Results are provided for all segments with significant representation in the LRA effort; these segments are organized in Table 7–2 by IOU and sorted on the overall assessment score. The overall assessment score is a value between three and minus three, with a maximum or minimum value of one for each issue area. In addition to that score, the table also includes the number of projects assessed and the number of the 17 issue areas determined to be both below average and above average. A more detailed examination of these issue areas is included below in Section 7.2.4 and in Appendix E, Section 2. Appendix E also includes a series of segment groupings that feature cross-program (or program group) comparisons. These additional tables provide an opportunity to better explore the reasons for below average performance and point to key issue areas for targeted improvement at the segment level. The relative performance of other programs or segment groupings can be examined by reviewing the tables with a focus on the significant findings indicated with dark and light shaded cell values, whereas the dark shading reflects below average performance and areas for improvement, and the light shading reflects above average performance.

PG&E LRA results show that the core calculated programs and the Savings by Design new construction offering outperform all other groups with relatively high scores and a relatively low number of below average assessments. PG&E REEP has strong performance while there is room for considerable improvement for PG&E CCC and government program offerings in general. The latter represent the 3 out of 19 PG&E segments that perform below the average overall assessment score overall assessment score for the complete portfolio sample. The findings from the LRA assessments show scores roughly in line with M&V results in Chapter 5. However, the government partnership programs had relatively good GRRs (but were not uniformly high).

LRA results for SCE show that ten out of twelve program groups perform above the average overall assessment score for the complete portfolio sample. The SCE program group performing the best is the SCE Agricultural Calculated Energy Efficiency Program and the poorest performing program in this group is the SCE Community Colleges of California (CCC) program. While the new construction programs are typically higher performing, the SCE new construction program scores near the bottom of the list of SCE program groups (although still better than the average overall assessment score. The gross impact results for the SCE new construction programs are average compared to overall gross impact results.

SCG results show that all of their program groups perform at or below the average overall assessment score for the complete portfolio sample. The program requiring the most attention is the SCG Commercial Calculated program. The lower rigor assessment process is very informative of the issues facing the SCG programs and attention to the noted issue areas can greatly improve performance. In contrast to the LRA scores, the SCG commercial core program was a better performer than the industrial program.

SDG&E results show that six out of seven program groups perform below the average overall assessment score for the complete portfolio sample. While the SDG&E electric domain performs the least favorably of this group, however, the SDG&E New Construction program has the highest score for this IOU and is higher than the overall average. The gross impact results for this IOU are very different than the LRA findings. Based upon unweighted GRR averages, the gross impact results were average or slightly higher than average for the SDG&E core and third party programs, but low for the new construction program.

Statewide program results show that two out of nine program groups perform below the average overall assessment score for the complete portfolio sample. The Community Colleges of California (CCC) program is the worst performing program group preceded by all third party programs, second-to-last of the statewide groups. The statewide group of new construction programs performs the best out of all the segments assessed, followed by UC/CSU at second best.

Table 7–2: Lower Rigor Assessment Results Overview

Program Group Assessment	Number of Assessments (N)	LRA Score[7]	Below Average Assessments		Above Average Assessments	
			Count	Percent (%)	Count	Percent (%)
All LRA Average	536	-0.50	n/a	n/a	n/a	n/a
Pacific Gas and Electric						
PG&E New Construction	13	1.83	1	5.9	11	64.7
PG&E Industrial Calculated Incentives	36	1.77	1	5.9	11	64.7
PG&E Core	104	1.13	1	5.9	7	41.2
PG&E REEP	14	1.03	1	5.9	7	41.2
PG&E Commercial Calculated Incentives	37	0.93	3	17.6	8	47.1
PG&E Agricultural Calculated Incentives	19	0.90	3	17.6	8	47.1
PG&E Gas	97	0.93	1	5.9	6	35.3
PG&E RCx	10	0.80	3	17.6	7	41.2
PG&E Electric	172	0.73	1	5.9	5	29.4
PG&E Energy Efficiency Services for Oil Production (Global)	20	0.63	4	23.5	8	47.1
PG&E	252	0.37	0	0.0	2	11.8
PG&E Heavy Industry	12	0.13	4	23.5	5	29.4
PG&E Pump Energy Efficiency Services	12	0.00	4	23.5	4	23.5
PG&E Non- Core	148	0.00	2	11.8	2	11.8
PG&E UC/CSU	12	-0.03	5	29.4	5	29.4

[7] See Appendix E for an explanation of the overall assessment score and how lower rigor assessments were used to determine relative significant performance on a qualitative basis.

Table 7–2: Lower Rigor Assessment Results Overview

Program Group Assessment	Number of Assessments (N)	LRA Score[7]	Below Average Assessments		Above Average Assessments	
			Count	Percent (%)	Count	Percent (%)
All LRA Average	536	-0.50	n/a	n/a	n/a	n/a
PG&E CCC	11	-0.40	5	29.4	3	17.6
PGE "Energy Watch"	13	-0.70	5	29.4	1	5.9
PGE "Energy Watch" + Rightlights Program	14	-0.87	7	41.2	2	11.8
PG&E Other 3P	28	-1.10	9	52.9	3	17.6
Southern California Edison						
SCE Industrial Calculated Energy Efficiency Program	15	1.67	1	5.9	10	58.8
SCE Nonmetallic Minerals and Products (3P)	11	0.80	2	11.8	7	41.2
SCE Commercial Calculated Incentives Program	20	0.40	4	23.5	6	35.3
SCE UC/CSU	18	0.40	2	11.8	4	23.5
SCE (SCE Electric)	139	0.33	1	5.9	3	17.6
SCE Core	53	0.03	4	23.5	4	23.5
SCE Non-Core	86	-0.03	2	11.8	2	11.8
SCE Other 3P	25	-0.03	4	23.5	4	23.5
SCE "Energy Leader"	10	-0.07	3	17.6	3	17.6
SCE New Construction	22	-0.13	5	29.4	4	23.5
SCE GP	29	-0.40	5	29.4	3	17.6
SCE CCC	13	-1.23	8	47.1	1	5.9
Southern California Gas Company						
SCG Industrial Calculated	40	-0.13	4	23.5	4	23.5
SCG & SDG&E Gas	72	-0.50	6	35.3	3	17.6
SCG	64	-0.67	6	35.3	2	11.8
SCG Core	62	-0.87	7	41.2	2	11.8
SCG Commercial Calculated	14	-1.07	8	47.1	2	11.8

Table 7–2: Lower Rigor Assessment Results Overview

Program Group Assessment	Number of Assessments (N)	LRA Score[7]	Below Average Assessments		Above Average Assessments	
			Count	Percent (%)	Count	Percent (%)
All LRA Average	536	-0.50	n/a	n/a	n/a	n/a
San Diego Gas & Electric[8]						
SDG&E New Construction	19	0.70	4	23.5	8	47.1
SDG&E Non-Core	50	-0.60	7	41.2	4	23.5
SDG&E Commercial Calculated	24	-0.67	7	41.2	3	17.6
SDG&E Core	31	-1.07	7	41.2	1	5.9
SDG&E BID	29	-1.33	9	52.9	2	11.8
SDG&E	81	-1.67	11	64.7	2	11.8
SDG&E Electric	73	-1.83	11	64.7	1	5.9
Statewide Program Groups						
All New Construction	56	0.73	2	11.8	6	35.3
All UC/CSU	30	0.57	1	5.9	4	23.5
All Core	249	0.03	1	5.9	1	5.9
LGP	23	-0.20	3	17.6	2	11.8
SGP	68	-0.37	5	29.4	3	17.6
DGS	10	-0.50	6	35.3	3	17.6
Third Party	139	-1.07	5	29.4	1	5.9
All CCC	24	-1.27	8	47.1	1	5.9

7.2.4 Results by IOU Fuel Domain

This section presents LRA results by IOU fuel domain. The sample sizes by utility vary with a low for the electric utilities of 73 for the SDG&E electric domain and a high of 172 for PG&E electric domain. For gas utilities, the sample size varies between 72 for SCG/SDG&E gas and 97 for PG&E gas. The highlighted cells in Table 7-3 below indicate “above average” (light grey) or “below average” (dark grey) performance for this grouping, as discussed in Section 7.2.1. For comparison purposes the average result across all sample points is shown in the furthest right column in the table.[9]

[8] The SDG&E gas domain is not included due to having fewer than ten valid LRA sample points.

[9] “All LRA Average” is the average score for completed assessments across all program groups.

Table 7-3: IOU Fuel Domain Program Performance, Percent "Poor"

Issue Assessed	IOU Fuel Domain					
	PG&E Electric	**PG&E Gas**	**SCE**	**SCG and SDG&E Gas**	**SDG&E Electric**	**All LRA Average**
Number of Assessments	172	97	139	72	73	536
Number of "above average" issues	5	6	3	3	1	–
Number of "below average" issues	1	1	1	6	11	–
Overall Assessment Score [10]	0.73	0.93	0.33	-0.50	-1.83	-0.50
Project Documentation and Specification						
IOU application documentation complete and accurate	26%	23%	30%	17%	36%	26%
IOU tracking data complete and accurate	13%	22%	20%	24%	25%	20%
Project utilized pre-installation M&V	20%	15%	24%	25%	30%	22%
Appropriate baseline	14%	15%	22%	24%	37%	21%
Early replacement claim: valid RUL / EUL approach used	20%	13%	16%	17%	29%	18%
Appropriate Calculation Method						
Appropriate impact calculation method	18%	14%	22%	21%	33%	21%
All relevant inputs considered	12%	19%	20%	21%	25%	18%
Adequate values for all inputs	20%	21%	21%	22%	18%	21%
Appropriate HVAC interactive effects calculation method	18%	11%	56%	Small Sample	33%	33%
Appropriate non-HVAC interactive effects calculation method	29%	12%	17%	36%	0%	21%
Project utilized post-installation M&V	23%	26%	33%	38%	40%	30%
Compliance with Program Rules						
Measures are IOU program eligible	3%	7%	1%	1%	3%	3%
Measures exceed code or industry standard practice	12%	11%	6%	22%	*10%*	11%
Multiple IOU fuel impacts properly accounted for	60%	42%	Small Sample	38%	70%	52%
If applicable, fuel switching supported with three prong test	Small Sample	57%	Zero Sample	Small Sample	100%	69%
Non-IOU fuel and ancillary impacts accounted for (cogen/waste heat recovery/ refinery gas, etc.)	84%	56%	88%	64%	71%	73%
Customer installation meets program rules	10%	14%	9%	17%	*19%*	13%

10 See Appendix E for a detailed discussion of the overall assessment score.

Table 7-3 above summarizes the assessment results for the IOU fuel domains.

- The PG&E gas sampling domain receives a high overall assessment score, as does PG&E electric. PG&E gas and PG&E electric programs are above average in six and five issue areas, respectively, and below average in only one area. PG&E electric programs are below average in *appropriate non-HVAC interactive effects calculation method* whereas PG&E gas programs fall short of the average in *measures are IOU program eligible*. Since PG&E contains the largest number of LRA samples, it is also informative to review the PG&E-specific results in comparison to smaller groups of PG&E programs as presented in Appendix E.
- SCE programs are above average in three areas in the *compliance with program rules* category, falling significantly short of the average only in *appropriate HVAC interactive effects calculation method*.
- For the SDG&E gas and SCG sampling domain there are six below average areas requiring targeted improvement. The SDG&E gas and SCG domain programs would benefit from a more broad focus on process improvements including *IOU tracking data complete and accurate*, *all relevant inputs considered*, *appropriate non-HVAC interactive effects calculation method, project utilized post-installation M&V*, *measures exceed code or industry standard practice,* and *customer installation meets program rules*.
- SDG&E electric programs demonstrate consistently below average performance in all areas related to the *project documentation and specification* category. Performance is also below average for three out of six issues of the *appropriate calculation method* category and in the *compliance with program rules* category.

7.2.5 Results by Program and Program Group

This section highlights some key observations about how programs and program groups compare to each other and against the statewide average on the 17 key issue areas. Table 7-4 summarizes the assessment results for some of PG&E's larger third party programs and for PG&E's Pump Energy Efficiency Services program.

Table 7-4: 3rd Party and PG&E Pump Efficiency Program Performance, Percent "Poor"

Issue Assessed	PG&E Heavy Industry	PG&E Global	PG&E REEP	PG&E Pump EE Services	All LRA Average
Number of assessments	12	20	14	12	536
Number of "above average" issues	5	8	7	4	–
Number of "below average" issues	4	4	1	4	–
Overall Assessment Score	0.13	0.63	1.03	0.00	-0.50
Project Documentation and Specification					
IOU application documentation complete and accurate	25%	45%	7%	25%	26%
IOU tracking data complete and accurate	25%	25%	21%	8%	20%
Project utilized pre-installation M&V	8%	35%	0%	0%	22%
Appropriate baseline	17%	15%	21%	25%	21%
Early replacement claim: valid RUL / EUL approach used	42%	10%	36%	42%	18%
Appropriate Calculation Method					
Appropriate impact calculation method	33%	10%	7%	17%	21%
All relevant inputs considered	8%	0%	14%	17%	18%
Adequate values for all inputs	0%	15%	7%	25%	21%
Appropriate HVAC interactive effects calculation method	Zero Sample	Zero Sample	Small Sample	Zero Sample	33%
Appropriate non-HVAC interactive effects calculation method	Small Sample	Zero Sample	Small Sample	Zero Sample	21%
Project utilized post-installation M&V	33%	15%	0%	0%	30%
Compliance with Program Rules					
Measures are IOU program eligible	0%	0%	0%	0%	3%
Measures exceed code or industry standard practice	0%	30%	0%	8%	11%
Multiple IOU fuel impacts properly accounted for	Small Sample	Zero Sample	Small Sample	Zero Sample	52%
If applicable, fuel switching supported with three prong test	Small Sample	Zero Sample	Small Sample	Zero Sample	69%
Non-IOU fuel and ancillary impacts of project properly accounted for (cogen/waste heat recovery/ refinery gas, etc.)	Small Sample	Small Sample	Small Sample	Small Sample	73%
Customer installation meets program rules	25%	5%	14%	33%	13%

Some observations on Table 7-4 regarding PG&E's large third party programs include:

- The PG&E Global program receives above average assessments in eight issue areas with a strong showing in all areas of the *appropriate calculation method* category, however; the program is below average in four areas with three of these issue areas showing the worst performance of all the program groups shown in Table 7-4. Targeting the *application documentation*, *project utilized pre-installation M&V,* and *measures exceed* *code or industry standard practice* issue areas will significantly improve the performance of the PG&E Global program.

- The PG&E Pump Energy Efficiency Services Program receives four above average and four below average assessment scores. *Pre*- and *post-installation M&V* are the program's most notable above average performance areas with zero poor assessments. However, appropriate baseline, *valid EUL/RUL approach*, *adequate values for all inputs*, and *customer installation meets program rules* are the program's below average performance areas. While the program specifically allows like-for-like replacements and requires on-site inspections, greater attention to the project details and using those data appropriately in calculation tools would improve this program.
- The PG&E REEP (Refinery Energy Efficiency Program) and the PG&E Global (Energy Efficiency Services for Oil Production) programs perform above average relative to others in this group of programs. For PG&E REEP, the assessment scores are above average in seven areas and are below average in only one issue area, specifically *valid RUL/EUL approach*. Both the PG&E Global and PG&E REEP programs have significantly above-average performance within the *appropriate impact calculation method* category, but both also show below-average performance in other areas. Where PG&E Global is below average in *IOU application documentation complete and accurate*, *pre-installation M&V,* and *measures exceed code or industry standard practice*, PG&E REEP is significantly above-average in these same areas.
- The PG&E Heavy Industry program assessment shows a roughly equal number of above average and below average issue areas. The program is below average in *IOU tracking data complete and accurate*, *valid RUL / EUL approach*, *appropriate impact calculation method* and, *customer installation meets program rules*.

Table 7-5 summarizes the assessment results for the nonresidential new construction (NC) programs run by PG&E, SCE and SDG&E. The SCG new construction program received too few assessments and therefore is not discussed here.

Table 7-5: New Construction Program Performance, Percent "Poor"

Issue Assessed	SCE NC[11]	PG&E NC	SDG&E NC	All LRA Average
Number of assessments	22	13	19	536
Number of "above average" issues	4	11	8	–
Number of "below average" issues	5	1	4	–
Overall Assessment Score	-0.13	1.83	0.70	-0.50
Project Documentation and Specification				
IOU application documentation complete and accurate	41%	8%	32%	26%
IOU tracking data complete and accurate	9%	0%	37%	20%
Project utilized pre-installation M&V [12]	9%	8%	16%	22%
Appropriate baseline	36%	0%	11%	21%
Early replacement claim: valid RUL / EUL approach used	5%	0%	5%	18%
Appropriate Calculation Method				
Appropriate impact calculation method	18%	0%	5%	21%
All relevant inputs considered	23%	8%	21%	18%
Adequate values for all inputs	27%	23%	21%	21%
Appropriate HVAC interactive effects calculation method	40%	Small Sample	11%	33%
Appropriate non-HVAC interactive effects calculation method	Small Sample	0%	0%	21%
Project utilized post-installation M&V	55%	15%	26%	30%
Compliance with Program Rules				
Measures are IOU program eligible	0%	8%	5%	3%
Measures exceed code or industry standard practice	9%	0%	0%	11%
Multiple IOU fuel impacts properly accounted for	Zero Sample	Zero Sample	Small Sample	52%
If applicable, fuel switching supported with three prong test	Zero Sample	Zero Sample	Zero Sample	69%
Non-IOU fuel and ancillary impacts of project properly accounted for (cogen/waste heat recovery/ refinery gas, etc.)	Zero Sample	Small Sample	20%	73%
Customer installation meets program rules	14%	0%	21%	13%

[11] The only IOU to submit a new construction program as a "program of interest" is SCE. The other new construction programs with a sufficient number of samples are also included here.

[12] Pre-installation M&V is applicable to new construction programs that include "capacity expansion" elements.

Observations on Table 7-5 surrounding new construction program offerings include the following:

- PG&E's nonresidential new construction program group is the highest ranking program according to its overall assessment score. All issue areas perform above average except for one below average issue area - *measures are IOU program eligible*. This is notable because eligibility is well-defined by the codes and standards and is an issue often easily identifiable with a desk review of project documentation.
- SCE's new construction program shows five below average issues, the most of any of the new construction offerings included in this assessment. Notable areas of concern for SCE's program include *IOU application documentation*, *appropriate baseline*, *all relevant inputs considered*, *adequate values for all inputs* and *post-installation M&V*. The poorest performing issue areas are *project documentation* and *post-installation M&V*. These poor scores reflect the practice whereby SCE does not complete post-installation true-up for new construction projects and instead will submit savings estimates based upon arbitrary code-compliance occupancy schedules rather than using the facility's expected or actual occupancy schedule. The relatively poor performance is apparently worsened by SCE's inclusion of production-related capacity expansion projects that would benefit from post-installation M&V. However, SCE's program has a good showing in four issue areas including *measures are IOU program eligible, IOU tracking data complete and accurate*, and two other areas under the *project documentation and specification* category.
- The SDG&E New Construction program posts above average results in eight issue areas. Four significant below average performance areas are found with *application documentation*, *IOU tracking data complete and accurate, all relevant inputs considered* and *customer installation meets program rules*, and represent areas needing targeted improvement. These issue areas can all be mitigated through an improved desk review or similar lower rigor efforts.

Table 7-6 summarizes the performance of four statewide (SW) partnership programs including the California Community College (CCC) and UC/CSU institutional partnership programs. Interestingly, the SCE UC/CSU group is the strongest performer of this group, and the SCE CCC group is the weakest performer.

Table 7-6: Statewide Educational Partnership Program Performance, Percent "Poor"

Issue Assessed	SCE CCC	SCE UC CSU	PG&E CCC	PG&E UC/CSU	All LRA Average
Number of assessments	13	18	11	12	536
Number of "above average" issues	1	4	3	5	–
Number of "below average" issues	8	2	5	5	–
Overall Assessment Score	-1.23	0.40	-0.40	-0.03	-0.50
Project Documentation and Specification					
IOU application documentation complete and accurate	46%	44%	36%	50%	26%
IOU tracking data complete and accurate	15%	6%	18%	25%	20%
Project utilized pre-installation M&V	31%	17%	36%	42%	22%
Appropriate baseline	15%	11%	36%	8%	21%
Early replacement claim: valid RUL / EUL approach used	31%	6%	9%	8%	18%
Appropriate Calculation Method					
Appropriate impact calculation method	23%	17%	27%	33%	21%
All relevant inputs considered	31%	17%	18%	8%	18%
Adequate values for all inputs	46%	22%	18%	25%	21%
Appropriate HVAC interactive effects calculation method	Small Sample	20%	Small Sample	Small Sample	33%
Appropriate non-HVAC interactive effects calculation method	Small Sample	Small Sample	Small Sample	Small Sample	21%
Project utilized post-installation M&V	46%	22%	36%	33%	30%
Compliance with Program Rules					
Measures are IOU Program Eligible	8%	6%	0%	0%	3%
Measures exceed code or industry standard practice	23%	11%	9%	8%	11%
Multiple IOU fuel impacts properly accounted for	Small Sample	Small Sample	Small Sample	Zero Sample	52%
If applicable, fuel switching supported with three prong test	Small Sample	Zero Sample	Zero Sample	Zero Sample	69%
Non-IOU fuel and ancillary impacts of project properly accounted for (cogen/waste heat recovery/ refinery gas, etc.)	Small Sample	Small Sample	Small Sample	Zero Sample	73%
Customer installation meets program rules	15%	11%	9%	8%	13%

Observations on the statewide educational partnership programs in Table 7-6 include:

- The PG&E UC/CSU and PG&E CCC Programs perform slightly above average overall but score very poorly with respect to *application documentation*, the use of *pre-installation M&V*, and *appropriate impact calculation method*. For the CCC program *appropriate baseline* selection and the use of *post-installation M&V* are below average, while for the UC/CSU program *IOU tracking data* and *adequate values for all inputs* are below average.

- The SCE CCC program has a very low overall performance with below average scores in eight issue areas, and it was the only program in this group which scored below average in *early replacement RUL-EUL*, *all relevant inputs considered*, and *measures exceed code or industry standard practice*. Based upon lower rigor assessment metrics, additional scrutiny of SCE CCC project documentation is needed to improve performance.
- The SCE UC/CSU program performs above average in four issue areas, with only two issue areas indicating targeted need for improvement. The *IOU application documentation complete and accurate* issue area shows the most need for improvement; although it should be noted that all four partnership programs were below average in this area. *Measure eligibility* was found to be another area for directed attention in this program. The SCE UC/CSU program group appears to out-perform the other programs in Table 7-6 through attention to *IOU tracking data complete and accurate*, *appropriate baseline*, *valid RUL/EUL approach*, and *post-installation M&V* issue areas.
- Overall observations are that performance for these four educational partnerships is strongest in the *compliance with program rules* category, although it should be noted that all four partnership programs were below average in the *IOU application documentation* area.

Assessment results for IOU-specific sampling domains are shown in Table 7-7. Results include SCE Local Government ("Energy Leader"), PG&E Local Government ("Energy Watch"), SCE other 3P and PG&E other 3P. These last two program groups include miscellaneous third party programs that were combined for sampling purposes.

Observations on Table 7-7 surrounding local government and miscellaneous third party program groups include the following:

- The PG&E Energy Watch program group scores significantly below average for two of six issues within the *compliance with program rules* category, with considerably below average performance in the *appropriate impact calculation method* issue area. The PG&E Energy Watch program shares with the PG&E other 3P and SCE "Energy Leader" programs significant deficiencies in *pre*- and *post-installation M&V* issue areas. Results also indicate a need for targeted improvement in two other issue areas including *measure eligibility* and use of *measures that exceed code or industry standard practice*.
- Scores are notably below average in *pre- and post-installation M&V* for all program groups in Table 7-7 other than SCE other 3P.

Table 7-7: Local Government and Other Third Party Program Performance, Percent "Poor"

Issue Assessed	SCE Other 3P	PG&E Other 3P	PG&E LG "Energy Watch"	SCE LG "Energy Leader"	All LRA Average
Number of assessments	25	28	13	10	536
Number of "above average" issues	4	3	1	3	–
Number of "below average" issues	4	9	5	3	–
Overall Assessment Score	-0.03	-1.10	-0.70	-0.07	-0.50
Project Documentation and Specification					
IOU application documentation complete and accurate	24%	39%	23%	40%	26%
IOU tracking data complete and accurate	32%	25%	15%	20%	20%
Project utilized pre-installation M&V	12%	39%	46%	30%	22%
Appropriate baseline	20%	14%	23%	20%	21%
Early replacement claim: Valid RUL / EUL approach used	24%	39%	23%	20%	18%
Appropriate Calculation Method					
Appropriate impact calculation method	12%	29%	31%	20%	21%
All relevant inputs considered	24%	14%	8%	10%	18%
Adequate values for all inputs	20%	25%	23%	20%	21%
Appropriate HVAC interactive effects calculation method	Small Sample	Small Sample	Zero Sample	Zero Sample	33%
Appropriate non-HVAC interactive effects calculation method	Small Sample	46%	Small Sample	0%	21%
Project utilized post-installation M&V	8%	36%	54%	40%	30%
Compliance with Program Rules					
Measures are IOU program eligible	8%	0%	8%	0%	3%
Measures exceed code or industry standard practice	8%	14%	15%	10%	11%
Multiple IOU fuel impacts properly accounted for	Small Sample	Small Sample	Zero Sample	Zero Sample	52%
If applicable, fuel switching supported with three prong test	Zero Sample	Small Sample	Zero Sample	Zero Sample	69%
Non-IOU fuel and ancillary impacts of project properly accounted for (cogen/waste heat recovery/ refinery gas, etc.)	Small Sample	86%	Small Sample	Small Sample	73%
Customer installation meets program rules	8%	7%	15%	0%	13%

- The PG&E other 3P overall score is very low and the group receives below average scores in nine issue areas. The project documentation processes for this program group warrant follow up. Mixed performance is seen with PG&E other 3P programs and are above average for documenting the *appropriate baseline* but below average with *measures exceed code or industry standard practice*. Other areas of significant concern include *adequate values for all inputs* and *appropriate non-HVAC interactive effects calculation method*. (See also Table E-4 in the LRA appendix.)

- The SCE Energy Leader group is above average for three issue areas including two issues areas within the *compliance with program rules* category and *all relevant inputs considered* issue area. The group has three areas of concern including below average scores in *application documentation* and in *pre- and post-installation M&V*.
- The SCE other 3P group received below average scores for four issue areas, most notably with the lowest score of this group of programs in the *all relevant inputs considered* issue area. This program group has relatively poor overall results with an equal number of above and below average issues areas and thus, warrants some concern.

Table 7-8: Statewide Institutional Partnership Program Performance, Percent "Poor"

Issue Assessed	UC/CSU	DGS	CCC	SGP	All LRA Average
Number of assessments	30	10	24	68	536
Number of "above average" issues	4	3	1	3	–
Number of "below average" issues	1	6	8	5	–
Overall Assessment Score	0.57	-0.50	-1.27	-0.37	-0.50
Project Documentation and Specification					
IOU application documentation complete and accurate	47%	20%	42%	40%	26%
IOU tracking data complete and accurate	13%	20%	17%	15%	20%
Project utilized pre-installation M&V	27%	50%	33%	34%	22%
Appropriate baseline	10%	30%	25%	19%	21%
Early replacement claim: valid RUL / EUL approach used	7%	10%	21%	13%	18%
Appropriate Calculation Method					
Appropriate impact calculation method	23%	30%	25%	26%	21%
All relevant inputs considered	13%	0%	25%	16%	18%
Adequate values for all inputs	23%	30%	33%	32%	21%
Appropriate HVAC interactive effects calculation method	29%	Small Sample	Small Sample	33%	33%
Appropriate non-HVAC interactive effects calculation method	Small Sample	Small Sample	0%	0%	21%
Project utilized post-installation M&V	27%	60%	42%	38%	30%
Compliance with Program Rules					
Measures are IOU program eligible	3%	10%	4%	4%	3%
Measures exceed code or industry standard practice	10%	10%	17%	13%	11%
Multiple IOU fuel impacts properly accounted for	Small Sample	Zero Sample	Small Sample	Small Sample	52%
If applicable, fuel switching supported with three prong test	Zero Sample	Zero Sample	Small Sample	Small Sample	69%
Non-IOU fuel and ancillary impacts of project properly accounted for (cogen/waste heat recovery/ refinery gas, etc.)	Small Sample	Zero Sample	Small Sample	50%	73%
Customer installation meets program rules	10%	10%	13%	12%	13%

Table 7-8 presents results for several statewide sampling domains including the four institutional or governmental partnership ("GP") groups: CCC,[13] UC/CSU,[14] SGP,[15] and DGS.[16] Only three assessments were completed for a CDCR (Department of Corrections) partnership, so this group is excluded from this analysis except as they are represented within other groups. Other observations on this table include the following:

- The UC/CSU group has above average performance in four issue areas, notably *appropriate baseline* selection. Still, there is one mentionable area for improvement, specifically *IOU application documentation.*
- The DGS group displays mixed performance overall with an overall assessment score equal to the LRA average. Performance is above average in three areas, notably including *project documentation* and *all relevant inputs considered.* DGS is the only program group in this table to perform above average in *project documentation* (the other three program groups are significantly below average) and zero "poor" assessments were found in the *all relevant inputs considered* issue area. There are six areas of considerable concern, including *pre- and post-installation M&V*, *appropriate baseline*, and *program eligibility* (with DGS being the worst performer on these issues within this group).
- The CCC partnership program exhibits below average performance with eight issue areas and only one above average area. Scores indicate four areas of concern in the *appropriate calculation method* category and several other issue areas including *project documentation, pre- and post-installation M&V*, and *measures exceeding code or industry standard practice*. Overall, the CCC programs display the greatest number of "poor" scores for this group of programs and the lowest overall LRA score.
- The SGP program group shows mixed performance with three areas above and five areas below average. Two of five issues in the *project documentation* category and three of six issues in the *appropriate calculation method* category are below average, while one area in each of the three categories shows above average performance.

7.3 LRA Overall Findings

Appendix E contains several tables in addition to those shown in the previous section that highlight other programs and groupings and which supplement the findings discussed in this chapter. The following general observations can be made with respect to performance within

[13] California Community College Partnership

[14] University of California and California State University Partnership

[15] State Government Partnership

[16] California Department of General Services Partnership

IOUs, within IOU fuel domains, and for similar programs across those domains, or on a statewide basis, drawing from findings discussed in this chapter and Appendix E.

- PG&E program documentation in the 2010-2012 cycle shows above average lower rigor assessment-based performance in the core, non-core, and new construction (NC) programs. The NC programs were below average only in ensuring *measures are program eligible*. The commercial core programs have room for improvement with three issues related to *compliance with program rules*, while the agricultural core programs had shortcomings in the *IOU tracking database, all relevant inputs considered,* and *adequate values for all inputs* issue areas. Below average performance is noted for PG&E's smaller third party and government partnership programs.
- As shown here and in Appendix E, Tables E-11 through E-14, SCE programs are slightly above average, in general. The most significant area of concern for SCE may be with the CCC programs. The new construction and core programs share one common area for concern in *considering all relevant inputs*. The SCE NC program has room for improvement in *IOU application documentation, appropriate baseline, all relevant inputs considered, adequate values for all inputs*, and a *lack of post-installation M&V* (specifically for capacity expansion projects). The SCE core program group has below average performance in the areas of *inappropriate impact calculation methods, relevant inputs not considered, inappropriate HVAC interactive effects calculation methods,* and not appropriately documenting *non-IOU fuel and ancillary impacts of projects*.
- For SCG, as shown in Appendix E, Table E-15, the commercial and industrial core programs (SCG3607 and SCG3611) share a significant lack of performance on *measures exceed code or industry standard practice.* Performance was generally below average among the SCG program groups with the SCG Commercial Calculated program performing the worst of SCG's program groups. The SCG Industrial Calculated program has below average performance in *tracking system documentation, appropriate non-HVAC interactive effects calculation method* and *compliance with program rules*.
- From Appendix E, Table E-16 we can see that the SDG&E programs were rated below average for most of its program groups and across many of the 17 issues. There are some similarities between the overall performance for SDG&E core and non-core programs with below average overall assessment scores and seven below average issues areas, but they share only two common areas for improvement in sufficient *pre-installation M&V* and *customer installation meets program rules*. However, as discussed in the appendix, the NC program (SDGE3118) is above average, with an overall assessment score slightly above the statewide new construction average. The issues of greatest concern for SDG&E new construction include the *tracking data, application documentation* and *compliance with all program rules*. The SDGE BID program has the lowest scores relative to other SDG&E results with nine below average issues areas. *Inappropriate*

baseline, inappropriate impact calculation methods, and *not considering all relevant inputs* are areas for improvement for the BID program and the commercial core program; the commercial core program also shows below average scores in *measures exceed code or industry standard practice* and *installation meeting program rules*.

- Lower rigor assessment performance for similar programs does not indicate consistently above or below average performance for different IOU or IOU fuel domains; however, PG&E's strong core program performance seems to dominate results in many areas.

7.4 LRA Updates Based on M&V Activities

This section discusses findings arising from a component of the lower rigor assessment effort initiated at the beginning of the AD Period – the LRA update process. Briefly, the lead evaluator for each project was responsible for updating the lower rigor assessment form with new information (after all supplemental data requests were fulfilled and gross impact activities were complete). This supports a comparison between the LRA update-based results and the initial desk review of IOU project documentation, and is the subject of this section of the report. For any M&V step that materially changed the prior assessment, a field was populated to indicate which M&V step caused the LRA assessment to change. The goal was to determine if the lower rigor assessment process is a good indicator of project performance and what additional evaluation activities, beyond the initial desk review, are needed to achieve greater levels of understanding and accuracy. In general, the results following the LRA update process are more robust and accurate because updates are based on additional gross impact-based information sources. The results are presented in Table 7-9 in terms of the frequency of changes to "good", "neutral", or "poor" for each issue area (regardless of the "reason for change") and in Table 7-10 in terms of the reasons for change.

The lower rigor assessment updates caused changes to 86 percent (or 203) of the 236 AD period projects covering 419 discrete assessments of project issues. Table 7-9 shows that the most common issue to be changed to "good" from "unable to assess," "poor," or "neutral" as a result of new information received after the initial desk review is *multiple IOU fuel impacts properly accounted for* (15 percent) but this also changed in 9 percent of the cases to "Poor".

Table 7-9: Changes to LRA Assessments by Assessment Issue

Key Issue Assessed	Number of Changes	Assessment Changed To:		
		Good	Neutral	Poor
Number of assessment changes	419	26	151	233
Project Documentation and Specification				
IOU application docs complete and accurate	24	4%	9%	4%
IOU tracking data complete and accurate	11	4%	3%	2%
Project utilized pre-installation M&V	14	8%	6%	1%
Appropriate baseline	53	12%	10%	15%
Early repl. claim: valid RUL / EUL approach	30	4%	8%	7%
Appropriate Calculation Method				
Appropriate impact calculation method	52	4%	14%	13%
All relevant inputs considered	38	8%	9%	10%
Adequate values for all inputs	32	8%	7%	8%
Appropriate HVAC interactive effects calc. method	3	4%	0%	1%
Appropriate non-HVAC int. effects calc. method	4	0%	1%	1%
Project utilized post-installation M&V	52	4%	14%	13%
Compliance with Program Rules				
Measures are IOU program eligible	36	12%	11%	7%
Measures exceed code or industry standard practice	20	4%	7%	3%
Multiple IOU fuel impacts properly accounted for	36	15%	8%	9%
Fuel switching supported with 3 prong test	8	4%	1%	2%
Non-IOU fuel & ancillary impacts accounted for (cogen/waste heat recovery/ refinery gas, etc.)	3	0%	0%	1%
Customer installation meets program rules	3	4%	0%	1%

The most common issue to be changed to “poor” from “unable to assess,” “good,” or “neutral” is *appropriate baseline* (15 percent), but initial *appropriate baseline* scores also changed to “good” in 12 percent of the cases. Finally, the two most common issues to be changed to “neutral” from “unable to assess,” “good,” or “poor” is *appropriate impact calculation method* and *project utilized post-installation M&V* (14 percent each), but each of these issue areas also changed in 13 percent of the cases to “Poor”.

These data illustrate that the lower rigor assessments change as a result of being updated after more rigorous M&V activities, as 86 percent of the sites were subject to some kind of change. However, the magnitude of these changes is not apparent. It is significant that changes were made for 10 to 15 percent of the cases in important areas that affect gross impacts, such as baseline, eligibility, and calculation methods, and these were also noted as being key discrepancy factors in the gross impact evaluation of M&V points.

Table 7-10: Changes to LRA Assessments by Reason for Change

Reason for Change	Number of Changes[17]	Percent of Changes	Assessment Changed To:		
			Good	Neutral	Poor
Number of assessment changes	**419**	**–**	**26**	**151**	**233**
Initial desk review	n/a	n/a	n/a	n/a	n/a
Desk review QA/QC	125	29.8%	42%	46%	19%
IOU data request	6	1.4%	4%	3%	0%
Customer recruitment contact	1	0.2%	0%	0%	0%
IOU account rep phone call	0	0.0%	0%	0%	0%
Total changes prior to on-site visit	**132**	**31.5%**	**46%**	**50%**	**19%**
NTG interview	0	0.0%	0%	0%	0%
First on-site visit	206	49.2%	46%	32%	62%
Second on-site visit	0	0.0%	0%	0%	0%
Customer follow-up contact	6	1.4%	0%	1%	2%
Savings analysis calculations	63	15.0%	8%	15%	16%
Final Site report drafting	2	0.5%	0%	1%	0%
Final Site report QA review	1	0.2%	0%	1%	0%
Total changes after on-site visit	**278**	**66.3%**	**54%**	**50%**	**81%**

Table 7-10 shows that 31.5 percent of changes to the lower rigor assessments happened prior to an on-site visit and the vast majority of those changes were due to additional desk review. A small number of changes resulted from additional IOU data requests. A full 66.3 percent of changes occurred after an on-site visit, with 59.2 percent of those changes resulting in a “poor” rating.

Of the 233 changes to a “poor” rating, 81 percent of those changes required information that was only available from an on-site visit. This suggests that 81 percent of the changes would otherwise be undetectable without a more rigorous M&V approach involving site visits and savings estimation. While the lower rigor assessment process is important in guiding the on-site data collection process, the LRA efforts alone do not necessarily affect the GRR. These findings provide motivation to augment lower rigor assessments, if used in the future, with on-site data collection activities.

7.5 LRA Process Findings and Recommendations

This section examines the LRA process and provides recommendations for the future use of LRA in evaluations and other program functions. The original intent of the LRA effort was to

[17] The number of changes reflects the total number of changes for all projects; some assessments (with changes to good, neutral, or poor) covered both electric and gas assessments, therefore; the number of assessment changes is slightly lower than the number of project changes.

expand the scope of the custom impact evaluation to include smaller programs that would otherwise not benefit from the M&V process.

It is anticipated that a process and form similar to the LRA will continue to be used in subsequent custom impact evaluation efforts as part of the normal engineering review of custom projects. Furthermore, future evaluation activities should consider the use of the related custom ex-ante review activity and form. This may consist of Project Practices Assessments (PPAs) similar to the LRAs which feature a review of project compliance with ex ante review guidance and requirements, and conformance with evaluation and policy guidance, as well as program rules, with an emphasis on ex ante gross impact development/methods and possibly free ridership assessment. It is also recommended that the IOUs integrate the use of these forms and activities in order to assist with internal engineering review and enforcement of program rules and CPUC guidance/policy.

The LRAs are useful for considering how effectively program rules are followed and how rigorously ex ante impact calculations are applied to the available project data. LRA is also effective for documenting project findings by specific issue area. The LRA process is very cost effective, fits squarely into the first steps of the detailed M&V process, and supplements and informs the on-site data collection plan. The development of the Excel-based form allowed easy compilation of results. If coupled with phone interviews and on-site activities, the LRA results can be even more insightful and more indicative of evaluation-based gross impact results.

One of the major areas for improvement is in the predictive power of the LRA efforts. The LRA score and the gross impact realization result were not strongly correlated. Additional attention may be beneficially directed at the algorithm for scoring, which previously used a simple, approximately equal weighting of issue areas. Evaluators and other stakeholders could develop additional scoring mechanisms which attempt to correlate LRA performance with GRRs. This might be based upon a regression analysis of the lower rigor assessment metrics, least squares analysis, differing weights for issues of critical importance, or other methods.

Itron recommends that the IOUs consider using a process similar to the lower rigor assessment to review project applications, screen project applications for further review and develop consistency and reliability of results across programs and implementers, especially where standardization is needed. Any entity applying the LRA process should consider robust training and other methods of standardizing assessments.

8

Detailed Evaluation Findings and Recommendations

In this chapter we present key findings, drawn from across the previous results chapters of this report, and associated recommendations. Findings and recommendations are organized into the following sections:

- 8.1 Gross impact-related findings and recommendations
- 8.2 Net-to-gross-related findings and recommendations
- 8.3 Program findings and recommendations based on LRA results

As summarized in Chapter 5, ex-post gross impact realization rates were low (generally around 0.6) for all IOU fuel domains but one as compared to unity and the 0.9 default GRR for calculated programs for the PY2010 – 2012 program cycle. Net-to-gross ratios were low to moderate, ranging from 0.45 to 0.56. At a summary level, the detailed recommendations in this chapter fall into the following primary areas:

- To address overestimation, on average, of ex ante savings estimates, the IOUs should:
  - Improve estimation and quality control of project operating conditions, ex-ante baseline determinations, calculation methods, and eligibility rules to address the discrepancy factors presented in this report, and
  - Improve adjustments to project savings based on post-installation inspections and M&V, where appropriate, while considering overall cost effectiveness constraints.
- To achieve sufficient quality control, increase consistency between project files and tracking data, and minimize miscommunication of project claims. IOU project documentation and tracking data needs to significantly improve.
- To reduce continued moderate to high free ridership, IOUs should test program features and procedures changes focused on increasing program-induced savings.

8.1 Gross Impact-Related Findings and Recommendations

Many of the identified gross impact issues are similar to areas highlighted for improvement in the *WO033 Interim Report* and in recent custom-related impact evaluations for previous program cycles.

As presented in Chapter 5, it was found that operating conditions, baseline specification, and calculation methods were all important discrepancy factors that contributed to impact-related differences between ex-post evaluation results and utility savings claims. Program improvements in these three areas alone could significantly improve the level of agreement between utility ex-ante and evaluation ex-post gross impact estimates.

Gross impact findings and recommendations are presented in the following subsections:

- Operating Conditions
- Baseline Specification
- Calculation Methods
- Cross-Cutting and Other Gross Impact-Related
- Documentation-Related

8.1.1 Operating Conditions

The most common discrepancy factor between claimed and evaluated savings was due to differences between ex-ante and ex-post estimates of operating conditions. This factor also accounted for a large percentage of the lower ex post evaluation results. Estimate operating conditions can be difficult in practice for a variety of reasons; however, some aspects of this estimation can be addressed through improvement in program implementation activities and quality control.

Finding: Changed Operating Conditions for Projects

Evaluated operations conditions were often found to be different than described in program project documentation. Sometimes the operating parameters found during the original analysis had changed due to production requirements, or longer term data collection had shown that project-documented operating conditions were changed or had existed only for a short period (E044, F220, H308). In other cases, it appears that project documents or tracking data (i.e., official IOU savings claims) were not updated to reflect information collected or available from the project site (E096). In a few projects, the measure had unexpected production effects or the measure had failed and had not been returned to operation (F001). Per evaluation guidelines, which dictate that the measures are to be evaluated as-found, the ex-post savings analyses were performed for the as-observed conditions, including back-casting where relevant to current operations, and did not include any forecasting or retroactive calculated impacts of measure operation that could not be verified.

- **Recommendation:** ***Increase focus on: a) accuracy of operating conditions, b) use of pre- and post-installation data and information, and c) keeping project documentation***

and tracking claims up to date with field information. The IOUs should ensure the use of realistic inputs and assumptions concerning operating conditions and take steps to address the tendency observed for PY2010 – 2012 to estimate operating conditions that lead, on average, to over-estimates of energy savings.

IOUs should consider increased use of, and improve incorporation of, data collection and monitoring to ensure a meaningful and accurate set of inputs or assumptions surrounding operations. Post-retrofit inspections should fully incorporate verification of measures, proper installation and operation, and any observed or otherwise known changes or deficiencies. IOU staff should check that pre-installation and post-installation reports are well organized and complete, with measure counts, changes in operation, efficiency values, and operating data, such as hours of use and supporting metering data, whenever possible.

Consideration should be given to selecting an appropriate and representative time period to use for data collection and savings determination. Increased use of selective measurement and short term monitoring is also recommended, subject to overall program cost-effectiveness constraints.

In some cases, IOUs should delay claiming energy savings for projects if the installation is not complete or if operations are very unstable or unrepresented of expected ex post conditions. The IOUs should also ensure that savings estimates are always updated in the project documentation and tracking systems when operation conditions are found to have significantly changed.

IOUs should ensure incorporation of needed aspects of pre- and post-installation review, as specifically related to operating conditions, into program manuals by addendum and in their next revision. IOUs should delineate expectations for post retrofit inspection paperwork and require inspectors to identify and collect and record pertinent parameters such as hours of use and kW measurements as well as quantities in both pre-installation and post- installation efforts. IOUs should consider holding multiple trainings, regularly (e.g., quarterly), with internal staff, implementers, and IOU technical reviewers, to ensure improvement and enhanced documentation. Examples of thorough, complete pre- and post-installation reports could be provided in order to set standards for acceptable data collection and reporting, and expand current practices beyond a cursory verification that new equipment was present at a given site.

Finding: Operating Conditions Not Based on Observed Conditions

New construction and industrial project ex-ante savings were sometimes based on future expected increased occupancy, production, or throughput (E031). In some cases, the baselines for data centers were inconsistent with PG&E's Data Center Baseline document. Assumptions regarding high levels of data center occupancy and associated load increased ex-ante savings estimates in some cases (E005, F203). In some new construction cases, default Title 24 schedules were used in compliance software rather than actual or expected conditions for the particular project (found for several new construction projects). Industrial projects sometimes included forecasts of future production levels.

- **Recommendation*: Follow CPUC guidance to use observed conditions for savings calculations.*** Savings should be calculated based on existing evaluation guidance when the savings claim is finalized. CPUC evaluation and program guidelines for PY2010 – 2012 require savings estimation based upon observed conditions, not using a load that is defaulted from compliance software or forecasted over the life of a facility.[1]

8.1.2 Baseline Specification

Improper baseline specification resulted in a substantial number of adjustments, though less common for gas projects, resulting in significant impacts to ex-ante savings claims for both electric and gas projects. These adjustments often arose from improperly classified projects (H310, E057), improper baseline operation (H317), or crediting new or replacement equipment with improved efficiencies when, in reality, the new equipment efficiency did not exceed industry standard practice (E246, F006, G013, H329).

Finding: Baseline Not Identified and Defaulted to Existing Equipment.

Despite previous evaluation recommendations and CPUC ex ante policy guidance, the IOUs all continued (for PY2010-2012 custom projects) to commonly calculate savings using existing equipment as the baseline, even for normal replacement projects and even after the CPUC's ex ante-related decision. Often normal replacement projects are incorrectly classified as early replacement projects and lack any documentation or effort to justify these early replacement claims (H310, H328). Early IOU intervention and increased attention is needed to clarify measure baseline definitions for many projects (E059, E200). Pre-existing equipment was often used as a default baseline for savings estimation, implying early retirement while ignoring industry standard practice (F201), code requirements (H302), customer-specific equipment maintenance practices (E205), equipment remaining useful life (RUL), and evidence of program influence on early replacement (E057). Baseline determinations and RUL estimates were not provided by the IOUs in their PY2010-2012 tracking systems or applications. In many cases,

[1] Future work by the CPUC, as part of planned PY2013-2014 nonresidential new construction evaluation activities, may focus on whether changes to CPUC guidelines are warranted.

savings were calculated relative to an in situ baseline and those savings were assumed to persist over the entire period of the effective useful life (EUL) of the new equipment (H310).

- **Recommendation**: ***Increase efforts to* e*nsure conformance with CPUC baseline policies and eliminate common practice of defaulting baselines to existing equipment.*** The IOUs should mount a concerted effort to adopt baseline specification practices in conformance with Decision 11-07-030 and CPUC policy. Conformance to these guidelines and accurate specification of project baseline type, such as early retirement, normal replacement, replace on burnout, system optimization, new construction, etc. would eliminate many of these issues. The IOUs should consider amending program rules to eliminate incentive eligibility for measures that are not more efficient than code or ISP (or what would otherwise be required to meet performance requirements). Careful consideration must be given to avoid regressive baselines (baselines that are less efficient than current operations), as in a boiler project (G418). If the efficiency of the pre-existing equipment is higher than the replacement equipment baseline, then the IOUs should select the pre-existing equipment as the baseline.
- **Recommendation**: ***Clearly identify project event in terms of natural replacement, replace on burnout, early replacement, new construction, etc., and set the appropriate baseline accordingly.*** Realistic baselines based on code, current industry standard practices, or pre-existing equipment (with an associated RUL) should be clearly identified, supported and documented. The IOUs should carefully review current codes and any code changes that affect the baseline selection. If a claim is made for program-induced early retirement of functioning equipment, claims should include documentation of the remaining useful life (RUL) of the equipment replaced and the baseline used for the post-RUL period. *Some evidence of the adoption of this recommendation has been seen in the ex-ante review process and in the IOUs' 2013 procedures manual.*

8.1.3 Calculation Methods

As summarized in Chapter 5, it was found that the ex-ante calculations for an array of projects were lacking in terms of the calculation method applied and incorporation of correct inputs to describe pre- and post-installation operating conditions (improvements to capturing operating conditions more accurately are discussed in Section 8.1.1. Recommendations to improve calculation methods and protocols are provided below.

Finding: Inadequate Impact Methods and Models

Inadequate or suboptimal methods, models, and inputs were found in roughly 20 percent of projects included in the evaluation gross impact sample (E012, E338, G030, H215, H302).

- **Recommendation:** ***Continue to review and improve impact methods and models through review of evaluation results, industry best practices, and collaboration with the CPUC's en ante review process.*** The IOUs and their subcontractors should review the methods and models used in this evaluation for projects that were identified as having inadequate ex ante calculation approaches. IOUs should continue to improve their modeling approaches through systematic review and assessment of approaches developed and used internally, by third parties, by professional organizations, and by programs in other jurisdictions. In addition, the IOUs should continue to work closely and collaboratively with the CPUC's ex ante review process to assess and agree on modeling approaches based on the results of ex post evaluation and ongoing ex ante review.
- **Recommendation*:*** ***Carefully review ex-ante savings claims, inputs, and calculation methods*.** Ex-ante savings estimates and calculation methods should be more thoroughly reviewed and approved by IOU technical staff prior to finalization of incentives and savings claims. These reviews by knowledgeable technical staff can help mitigate issues caused by ineligible equipment, measures that are repair or merely code-compliant, and standard practice considerations (E209, F009, G038, H209).

Finding: Fuel Switching Not Properly Analyzed

Several sampled projects, primarily submitted by PG&E and SCG, involved fuel switching, cogeneration or distributed generation. These projects were not properly identified as involving fuel switching and often did not include a three-prong test and corresponding analysis and documentation as required by CPUC policy (E004, E057, E059). Due to the complexity of these projects, which usually occur in the industrial sector, it is often difficult for evaluators to establish fuel switching in cases where it is not correctly indicated in IOU documentation and claims.

- **Recommendation:** ***Identify fuel switching projects in claims and carry out three-prong test.*** Projects that include fuel switching effects should be clearly identified as such in utility tracking systems and include the three-prong test in accordance with CPUC policy.

Finding: Improper Reporting of Multiple Fuels and Interactive Effects

Many projects involved impacts on multiple fuels, but the savings claim and ex-ante analyses did not include identification of all of the affected fuels and estimate all of the associated impacts. Some projects claimed savings in one fuel (usually electricity) but neglected changes caused by the project in other fuels, which in some cases included increases in use (E209, H008, H402).

- **Recommendation:** ***Identify and calculate impacts on all affected fuels.*** All program-induced changes in fuel use should be included in savings claims and associated analyses including interactive effects.

Finding: Failure to Consider Grid Impacts Before Finalizing On-Site Generation-Related Savings Claim

In a number projects with on-site cogeneration, the IOU did not consider the impacts of on-site cogeneration appropriately when estimating the savings. In one case, the customer was only using IOU-provided power for three months in a year, but full credit was given in savings calculations for incentive determinations, including savings of non-IOU power (H202).

- **Recommendation:** ***Improve QC on cogeneration-related projects and offer incentives only for grid savings in conformance with CPUC rules***. The IOUs should evaluate the impact of efficiency projects served by cogeneration systems appropriately and only offer credit for the electric savings that actually occur on the electric grid. If the source gas for a cogeneration system is supplied using an IOU rate schedule that includes the PPP surcharge, and a given project results in a decrease in grid-purchased cogeneration gas use, then gas savings should be claimed. If the claim includes peak demand reduction, the interval billing data should be used to verify that the facility uses grid purchased electricity during the peak demand period and that the claimed savings is associated with reduction to those grid purchases.

Finding: Building Modeling Techniques and Tools Need Attention

Significant challenges exist in using building modeling for whole building analysis. There are many variables in these models, and for new construction projects in particular, a well calibrated model based on energy bills and interval data aids in reducing discrepancies. Default occupancy and usage schedules - the compliance run - were often used instead of schedules associated with the actual expected occupancy of the facility (H317). In some cases, simulation tools that weren't appropriate for assessing the installed measures were used for the ex-ante analysis.

- **Recommendation:** ***The IOUs should work with CPUC to define and promulgate appropriate impact estimation approaches with respect to whole building modeling and new construction***.

8.1.4 Cross-Cutting and Other Gross Impact-Related

This section presents gross impact-related findings and recommendations that cut across the three discrepancy factors discussed above or are unrelated to those particular factors but are still relevant to the gross impact aspects of this evaluation.

<u>Finding: Gross Impact Results for Common Measure Groups Indicate Uneven and Sometimes Poor Performance.</u>

The GRR performance of measure groups in the IOU fuel domains varied considerably but variation was not often dominated by a single or clear set of factors. The group performance often was significantly affected by one or two extreme cases, due partly to small evaluation sample sizes for many of the measure groups, and partly due to the magnitude of some of the extreme GRR cases. For example, VSD projects appeared to do very well for PG&E, but this was due to one extreme case with a very high realization rate. Heat recovery and boiler projects generally showed results that were not as dispersed as other groups, with operating conditions being the primary discrepancy factor for these projects and for VSDs. Compressed air projects were found to be poor performers for PG&E and SCE due, principally, to calculation methods and a large number of projects with zero savings. The HVAC and HVAC controls groups were noted to have widely dispersed GRRs for therms, kWh and kW savings. For those groups, the GRRs for kW often did not align with the kWh GRRs, due in many cases to improperly identifying peak period demand savings. Calculation methods and operating conditions were primary causes of discrepancies for this group.

- **Recommendation**: ***Increase attention on compressed air, HVAC, and HVAC control projects.*** The IOUs should pay particular attention to compressed air, HVAC, and HVAC control projects, focusing attention on calculation methods and incorporating the correct operating conditions.

<u>Finding: Gross Impact Results and Practice for Programs Offered by IOUs Varied Considerably</u>

As was observed at the measure group level, program group results varied considerably and, again, GRR variation was not often dominated by a single or clear set of factors. The program group GRR performance often was affected by one or two extreme outliers, due partly to small evaluation sample sizes for many of the program groups, and partly due to the magnitude of some of the extreme GRR cases. For instance, new construction projects appeared to perform well on average, but in fact, project-level gross impact performance was widely scattered. Core projects performed relatively well if one excludes the most extreme results in smaller programs, and the commercial core projects typically performed best in the IOU fuel domains studied (with the exception of PG&E gas, where the industrial program had better performance). Third party programs showed average GRR performance, neither very good nor very poor. A trend noted in this grouping for PG&E is that larger third party programs seem to substantially outperform smaller programs. Operating conditions was a common reason for GRR discrepancies in the third party programs.

In addition to the GRR analyses, the LRA results in Chapter 7 and Appendix E provide program group specific analyses based on larger samples than the GRR analyses but with less rigor.

- **Recommendation:** ***Review program-level GRR and LRA results and compare to internal IOU program performance assessments.*** The LRA results provide a set of directional findings that IOU program administrators should review in order to make their own judgment of whether the evaluation results generally align or contradict their own assessments of individual program impact related practices. Based on this assessment, the IOUs should provide feedback and corrective action and direction, as necessary and appropriate, to individual programs.
- **Recommendation:** ***Conduct periodic due diligence to ensure programs adhere to IOU and CPUC impact estimation policies, guidelines, and best practices*. *Continue to work closely and collaboratively with the CPUC's Ex Ante Review Process*.** Given the multitude of non-utility and utility programs**,** the IOU program administrators should consider interventions such as increased training and project scrutiny to ensure the most accurate savings claims consistent with eligibility, baseline and program rules. In addition, the IOUs should continue to work collaboratively with the CPUC's ex ante review process and look for ways to leverage lessons learned from that process to implement their own internal ex ante review of internal and third party programs.

Finding: Ineligible Projects were Claimed.

Many projects were found to be ineligible (E092, F018, G013, H029, H220), resulting in zero or low realization rates for those projects.

- **Recommendation:** ***Improve IOU program requirements, manuals, training, and quality control procedures in order to screen out ineligible projects***. A more thorough IOU review of ex-ante documentation for eligibility and program rules is needed. Screening of routine maintenance and repair measures is a pressing issue in need of attention. Several measures did not fulfill a program-required five year minimum life. This issue affected server virtualization projects claimed by SDG&E, oil well pumps in harsh environments, and so on.

Finding: Inadequate Calculation Methods and Analysis Periods for Retrocommissioning Projects

Inappropriate calculation methods were used to estimate the savings for MBCx (monitoring-based retro-commissioning) projects. IOUs used IPMVP Option C to estimate the savings with as few as three months of pre-installation and three months of post-installation usage data (E053, F205).

- **Recommendation**: ***Extend M&V period and increase use of appropriate engineering analyses for MBCx***. Until such time as IOUs can develop better savings calculation methodologies in concert with CPUC for these types of projects, the IOUs should, when proceeding with MBCx projects, base ex-ante savings on an appropriate M&V period and selected engineering analysis of measures implemented. Extra attention is needed to identify and apply an appropriate M&V approach.

Finding: EC Motors (ECMs) had Results with a Large Spread of GRRs.

EC motors for refrigeration evaporator fans and their associated fan controllers showed widely varying evaluation results (E111, E340, F064, H045). Calculation methods were noted as the primary reason for discrepancy. Measure count was also an issue.

- **Recommendation**: ***IOUs, in collaboration with the CPUC, should develop a standard calculation methodology that incorporates climate zone***. Implementation should include clear identification of motors replaced and controllers installed to resolve measure count discrepancies.

8.1.5 Documentation, Data Request, and Site Access-Related

Evaluation and CPUC policy and funding oversight are dependent on timely and complete project and tracking system documentation for a variety of functions including monitoring progress against goals, characterizing program populations, conducting ex ante reviews, assigning measure groups, developing sampling domains, implementing samples, conducting initial assessments of projects and making labor and resource assignments, preparing for and conducting detailed on-site M&V, understanding IOU project-related activities and decisions, completing ex post evaluation-based impact estimates, and comparing ex ante and ex post results. Evaluations need proper and complete project documentation, received in a timely manner prior to the start of on-site data collection activities. This is essential because it provides the base, starting point source of information for every project in the gross impact sample. Lack of documentation is a long standing issue identified in every previous custom evaluation conducted in California; nevertheless, despite repeated attempts to remedy this problem, responses to evaluation data requests continued to be a problem during evaluation of the 2010-2012 custom projects. This section discusses specific issues and recommendations related to project documentation, data requests, and site access for evaluation.

Finding: Poor IOU Tracking Data

Discrepancies continue to occur between the IOUs' tracking data and the reported information in the IOU documentation. For instance, for a number of project tracking data entries, project descriptions are poor and savings are not the final savings in the application documentation (H401). For example, in some cases, SDG&E discounted project savings in the tracking system by 40 percent versus what was shown in project documentation (H209), and this discount was not described or flagged in the tracking system. We were eventually able to determine that this practice was due to partial incentive payments, but considerable effort was needed determine this.

- **Recommendation: *Ensure consistency between tracking data and project files.*** Project documentation should match all data in the tracking system. IOU staff should thoroughly check documentation and cross-verify the tracking system entries with the program documentation before filing savings claims and sending files to evaluators. Project documentation should be thoroughly checked and should be cross verified with the reported tracking number before sending the dataset to the CPUC or evaluators. IOUs should formulate guidelines for and standardize all aspects of project tracking in order to meet all CPUC expectations with respect to accuracy and completeness. This should be done in order to ensure best practices in tracking of all pertinent project information, including savings and claims, and to ensure that data request responses fulfill all CPUC objectives.

Finding: Incomplete Project Descriptions and Analysis Documentation

The documentation for many of the sample projects was insufficient to initiate an appropriate independent analysis and investigation. Common deficiencies included the following:

- Project descriptions and measure names in the tracking system did not always identify the energy efficiency measure. Sometimes only a customer name was entered in these fields. In other instances, generic names such as “process” were used.
- Project descriptions in the application files were generally adequate; failures were observed in some smaller third party programs and government partnerships.
- Changes made throughout the lifetime of the project (from application to final inspections and IOU reporting) were often not incorporated in the tracking system, leading to confusion related to the final estimates and their supporting documentation.
- Savings calculations for projects, measures and sub-measures were often not provided.
- Calculations were provided via paper documents, locked spreadsheets, scanned documents, or spreadsheets with dynamic functions stripped out and replaced with static values.
- Pre-installation and post-installation inspection records are sometimes not provided or are cursory when provided.

  - When incremental costs were the appropriate basis for maximum incentive determination, in many cases, these costs were not clearly provided for review; only total project costs were provided.
  - Project baseline classification (early replacement, normal replacement, replace on burnout, new construction, capacity expansion, system optimization and add-on measures) was not clearly indicated in the tracking system or in the project documentation (except in very few cases where the measure required code compliance, as was the case with new construction or where the equipment was cited by the IOU to be past its useful life).
  - RULs were not listed for projects in the tracking system or in the documentation.
- **Recommendation:** ***Develop a final "closeout" report for each claim.*** We suggest IOUs develop a "final project closeout report" to ensure that all program forms, files and data points are accounted for and properly stored for later retrieval. The report should clearly identify what documentation is available for a project (and also what documentation is not available).

Finding: Poor Project and Measure Identification in Tracking Systems

Not all projects, measures and sub-measures were clearly identified in the project documentation and in the tracking system. Often, the tracking systems contained only customer names or buildings as the project description. This was especially evident in new construction projects using building simulation models, such as Savings by Design projects, and retro-commissioning / MBCx projects.

- **Recommendation:** ***Clearly identify all measures in project documentation files and tracking systems.*** As part of post inspection closeout, the IOU or third party implementer, or both, should make certain that all measures are clearly identified in the post-installation inspection, the tracking system, and final reports. Project descriptions could easily be enhanced with more description of the technology deployed and quantities of measures installed. The use of standard project descriptions across all IOUs would be beneficial.

Finding: Incomplete Project Documentation Supplied on Initial Data Request

Project documentation received from the IOUs in response to initial data requests was often not complete, failed to clearly describe the project scope (as noted above), and was hard to follow. In the course of the evaluation, program participants were generally able to provide much more thorough project documentation. Occasionally, when incomplete project documentation was originally provided, the same incomplete project documentation was provided again when additional data requests were made.

- **Recommendation:** ***Ensure data request responses are responsive, organized, clear and internally consistent.*** We recommend that the IOUs provide detailed, complete documentation for each project, such that the evaluator can see a clear trail from the project documentation to the tracking system-based savings estimates. The IOUs should assign knowledgeable staff to review calculations for accuracy and completeness. Staff should catalogue data for each project electronically in one location. IOUs should provide all data completely when needed for evaluation purposes. When further data requests are necessary, the evaluators recommend that much more detailed attention be given starting with a phone call between the IOU personnel assigned to provide the data and the lead evaluator for the site. Data responses from the IOUs should indicate clearly when the data requested is not in the possession of the utilities, their implementers, or other parties under IOU management, and whom to contact to obtain this information, if possible.

Finding: Standardize on Electronic Data

Project documentation provided by the IOUs to evaluators was composed of a diverse mix of electronic and hard copy data. Sometimes data were provided to evaluators in inappropriate formats, such as metering data in a PDF format. We do want to acknowledge that, *in an improvement from past cycles, more project documentation was available in electronic format than in previous program cycles. For instance, PG&E provided much more electronic data than in previous cycles.*

- **Recommendation:** ***As a general guideline, all project documentation should be compiled in one electronic location***. Formats for data storage should optimize the usefulness of the data source to a variety of potential users to ensure usability, especially of data and calculations. Passwords and existing IOU – CPUC security procedures can be utilized to ensure confidentiality of data. This approach supports ease of transfer and use, as well as addressing both environmental and security concerns.

Finding: Slow Response to Data Requests

IOU responses to data requests were often slow, and extensions in data request due dates were often requested. These delays made it difficult to provide evaluation results back to the IOUs and the public in a timely fashion.

- **Recommendation:** ***Prepare in advance for data requests to increase timeliness of response.*** IOUs need to develop plans to effectively fulfill data requests; aspects of this plan should include notification of coming data requests, personnel needs, electronic data retention policies, and partial fulfillment strategies. Extensions to data requests and incomplete fulfillment affect timely evaluation results. Custom project files should be maintained in a timely, proactive way that assumes they will be delivered to the CPUC and its contractors for evaluation purposes at any time, not in mode that is reactive to a data request.

Finding: Poor Organization of Project Documentation

Project documentation is often not well organized and is not consistent across projects, even within a given program. This lack of organization makes it difficult to efficiently review the documentation for purposes of setting up the evaluation and wastes time and resources.

- **Recommendation:** ***Set up standard templates for project documentation and enforce compliance with these templates.*** A table of contents should be provided to show where key data elements reside within the documentation.

Finding: Correct Building Simulation Models Sometimes not Provided

Evaluators often could not replicate ex-ante savings due to incomplete or inconsistent project data and models. In a number of projects, simulation models such as eQuest or Energy-Pro were used to develop ex-ante savings (E080, E086, E096, and E103). For some of these projects, the models that were provided as part of project analysis and documentation could not be rerun in a way that returned the same savings estimates that were included in the project files or the tracking data. This lapse caused unnecessary efforts to obtain the correct modeling files or rebuild models, increasing evaluation costs.

- **Recommendation:** ***Provide the final version of energy models and clearly identify the version of any simulation tool.*** This will increase the likelihood that models can be run with the appropriate version of the modeling tool and generate an equivalent result to the ex-ante savings claim to ensure a consistent starting point for the evaluation activity.

Finding: Site Access Issues Affected Integrity, Speed and Cost of Evaluation

In some cases it was difficult to obtain permission to access facilities in the M&V sample in order to perform site visits. In some instances, multiple attempts to schedule a site visit were rejected by customers who were reluctant to allow evaluators on-site. Some customers stated that the IOU had already verified the installed measures and they had already received the incentives, so they felt no need of further verification. Customer refusal of evaluation activities could lead to biased results, and it is hypothesized that customers with poorer performing projects might be more likely to refuse access to evaluation staff.

- **Recommendation:** ***Strengthen language in customer agreements making it mandatory for the customer, if selected, to participate in the CPUC's evaluation activities,*** including site visit(s), monitoring, data collection, telephone interviews, and the provision of customer collected data (such as production records). Account executives and implementers should reinforce with participants the potential for evaluation activities and likely timing. Current application language requiring cooperation varies across programs, but generally requires only that customers "cooperate" with evaluation activities.

8.2 Net-to-Gross / Program Influence Issues

This section presents findings and recommendations related to net-to-gross and program influence. Detailed NTG evaluation results are presented in Section 6 and Appendix D of this report.

<u>Finding: Free Ridership for Custom Projects Remains High Overall</u>

On a statewide basis, the NTGR across all program categories averaged 0.48 for electric programs and 0.53 for gas programs. These values indicate a medium high[2] level of free ridership, and a resulting medium low level of program influence, and are similar in magnitude to NTGRs from the past several evaluation cycles, as shown in Table 8-1. The general conclusions are that free ridership remains high for custom programs and further, that few adjustments appear to have been made with respect to either the custom program designs or their implementation procedures in order to reduce free ridership. While we are sensitive to the fact that it is not easy to provide the level of expertise needed at the right time to move industrial customers to higher levels of efficiency given their complex production- and site-specific processes, we also observe that very few readily identifiable steps have been taken by the programs with the specific goal of reducing free ridership.

[2] Medium high free ridership is defined in this report as between 50 percent and 74 percent (i.e., NTGR of between 0.26 and 0.50).

Table 8-1: Statewide Industrial Custom Program[3] Evaluation Net to Gross Ratios, Program Years 1998-2008

(1 – Free Ridership)	1998	1999	2000	2001	2002	2004-2005	PY2006-2008	
							PG&E	SCE
Weighted	0.53*	0.51	0.41	0.65	0.45	0.57	Electric - 0.45, Gas - 0.31	0.63

* Weighted by incentives rather than by kWh savings.

Some improvement over PY2006-2008 results was seen for certain program groupings (PG&E's Gas programs as an example). However, several other program groupings or categories experienced significant declines.[4]

Program influence was low in many cases for a number of different reasons. In some cases, program claims were made on a number of projects that customers initiated for non-energy savings reasons and for which no alternative was ever considered. In some instances program incentives were offered for measures and technologies that are industry standard practice (thus significantly increasing the odds of free ridership in any given application). Program attribution was also limited when program incentives were offered for projects that were being implemented by end users in response to mandates from other regulatory agencies (for example, citations from air resource districts). Further, for those projects already at an advanced stage, where equipment had already been budgeted and in some cases, orders already placed, program influence was very low. There were also instances where incentives were provided to firms that were already very advanced in their adoptions of energy efficiency, such as companies with established energy efficiency procurement policies or mandates, including national chain and big box stores.

- **Recommendation:** ***Adopt procedures to identify and affect projects with low program influence.*** The IOUs should carefully review projects during the project development stage for potential issues associated with a high likelihood of very low program influence. This process should provide timely feedback to program implementers regarding the estimated level of program influence. This would afford implementers an opportunity to influence projects found to have low program attribution by encouraging project decision makers to adjust the project scope to higher efficiency levels, where warranted.

---

3 From 1998 to 2005, the Standard Performance Contracting (SPC) program results are represented. The PY2006-2008 results are for the PG&E Fabrication, Process and Manufacturing Contract Group and the SCE Industrial Contract Group, respectively.

4 For example, the NTGRs for the SCE electric program domain have dropped from 0.631 in PY2006-2008 to 0.49 in this PY2010-2012 evaluation.

- **Recommendation:** ***Adjust the set of technologies that are eligible for incentives.*** Periodically review the list of qualifying measures for each program and eliminate eligibility for those that are standard practice. At a minimum, such reviews should take place annually. Measures that are already likely or very likely to be installed by a significant fraction of the market should, in most cases, not qualify for incentives. Although identification of such measures can be difficult in practice in the industrial sector, a number of such measures can be identified through investigation of industry practices (for example, interviews with manufacturers, distributors, retailers, and designers), analysis of sales data, and review of evaluation results. In determining which measures to retain and which to eliminate, a balance must be struck between reducing free ridership and avoiding significant lost opportunities. Ideally, sub-technology niche markets can be selected for the program that are less well established, but where substantial technical potential still lies.

  In addition, program implementers should *actively highlight and promote technologies that are less well-adopted, cutting edge, or emerging technologies*. Such measures are much less likely to be prone to high free ridership.

  Related, the *designation of the proper baseline* for a given measure type is critical. Program implementers should take great care in establishing program baselines and in developing a firm understanding of the underlying economics that most customers face when a given technology is acquired.

  For technologies that are already well established, another strategy is to *incent based on bundling of mandatory requirements or optional features* that enhance performance of the base technology. For example, VFDs can be required for premium efficiency motor and HVAC projects.

  Another option is to *use a comprehensive rather than a prescriptive approach* to discourage free ridership. For example, for water-wastewater plants, implementing a comprehensive approach and requiring the project to reach a minimum savings threshold (such as 15%) is less likely to be prone to high free ridership than a measure-level approach.

- **Recommendation:** ***Adopt procedures to limit or exclude known free riders.*** One way to accomplish this is to conduct screening for high free ridership on a project-by-project basis. In cases where likely high free ridership is found, the program implementer should encourage such customers to move to a higher level of efficiency or encourage a bundled retrofit to ensure deeper savings. Either of these options could result in funding a project that would not have been implemented absent the program. Another option is for the program to set the threshold for incentive eligibility higher across-the-board so that all such projects will need to meet a higher efficiency threshold to qualify.

One way to assess the rate of free ridership likely on a given project is to critically examine the key reasons behind the project **before** the incentive is approved. For example:

- Has the project already been included in the capital or operating budget? Has the equipment already been ordered or installed?
- Is the measure one that the company or other comparable companies in the same industry/segment routinely installs as a standard practice? Is the measure installed in other locations, without co-funding by incentives? Is the measure potentially ISP?
- Is the project being done primarily, or in part, to comply with regulatory mandates (such as environmental regulations)?
- Are the project economics already compelling without incentives? Is the rebate large enough as a share of incremental costs to make a difference in whether or not the project is implemented?
- Is the company in a market segment that is ahead of the curve on energy efficiency technology installations? Is it part of a national chain that already has a mandate to install the proposed technology?
- Does the proposed measure have substantial non-energy benefits? Is it largely being considered for non-energy reasons (such as improved quality, reduced labor costs, or increased production)?
- Is there a fungible efficiency element of the project, that is, is the equipment available only at a single bundled efficiency level, e.g., as could be the case with a highly specialized piece of process equipment? Related to this, if efficiency level is a malleable attribute of the project, were the costs and benefits of different levels of efficiency considered and quantified?

By conducting a brief interview regarding these issues before the incentive is approved, the implementer can better assess the likely degree of free ridership and may be able to then decide if the project should be excluded or substantially re-scoped to a higher efficiency level.

- **Recommendation:** ***Make changes to the incentive design.*** *Tier incentives by technology class*, such as equipment type, to enhance promotion of technologies that are less well accepted versus those that are already established. Under this approach, the incentive level for less widely adopted and emerging technologies would be higher, while the incentive level for more widely-adopted measures would be lower.

Consider Incorporating a Payback Floor, Excluding Projects for Which the Payback Time is Less Than, Say, One Year. Project-specific investigation of free ridership for custom programs also indicates that projects with extremely short payback periods are more likely to be free riders, all else being equal. Although it is certainly true that many customers do not adopt attractive efficiency projects with very low paybacks,[5] a payback floor can still be helpful, particularly if it is not set too high and if the administrator is allowed some flexibility in its application. Several program administrators in other parts of the country have used payback floors effectively, although such criteria present project cost verification challenges. A one year floor guideline makes sense because projects with a one-year payback or less can usually be funded out of the current year's energy budget. The use of a payback floor (a minimum payback level based on energy savings alone) can help to reduce free ridership by eliminating projects that have extremely quick paybacks and thus little need for ratepayer-funded incentives.

Offer bonuses to incent desirable behavior, such as installation of multiple measures or installation by a first-time participant.

- **Recommendation:** ***Use a sophisticated program design reflecting a comprehensive mix of program features and leveraging an array of delivery channels.*** Use a broad mix of program features and delivery channels to market projects. In addition to incentives, give full consideration, for example, to education and marketing outreach opportunities, technical/design services, upstream incentives in the technology manufacturing and delivery chain, commissioning of advanced systems, and other relevant intervention and delivery strategies. Conduct market research and convene focus groups to identify and test these concepts prior to program launch.
- **Recommendation:** ***More information is needed on industrial project costs, non-energy costs and benefits, net present value analysis, and associated participant cost-effectiveness analysis.*** There has been very little analysis conducted supporting the actual incremental cost of industrial and custom energy efficiency projects and further research is needed in this area. Rules of thumb, such as assuming that incentives represent half of incremental costs, appear to have been used instead as proxies. There is inadequate financial analysis conducted to determine what portion of the customer's financial investment threshold is associated with the energy savings of particular projects versus non-energy factors[6] such as increases in production and reductions in labor,

[5] For example, industrial end users sometimes do not invest in compressed air projects with paybacks as low as one year or even less.

[6] In custom projects, non-energy factors can sometimes drive project installation more than the economics associated with direct energy savings. Whether or not those factors also correlate with free ridership is likely related to the extent to which the program did or did not influence the end users' or trade allies' awareness, knowledge, and certainty of those benefits.

materials, and regulatory compliance costs. Increased financial analysis should be considered for inclusion in industrial project applications, especially for the projects with the largest incentives. A key reason for scrutinizing large incentive projects more fully is that the sheer size of such projects merits additional analyses as part of the project justification. Increased review of project financials inclusive of non-energy factors can also help to reduce free ridership.

8.3 Program Findings and Recommendations Based on LRA Results

Findings and recommendations stemming from the lower rigor assessment activities are presented in this section. Observations about issues that are significant for each IOU are described below.

Common issues existed across the entire LRA sample. For instance, the use of pre- and post-installation M&V was not a requirement for all programs and lacked uniformity of implementation across the portfolio of custom projects, but enhanced implementation of M&V would be expected to greatly improve the accuracy of gross impact savings estimates. It is an area with much opportunity for improvement. It is acknowledged by the evaluation team that the extent of M&V must be balanced against cost, targeted accuracy, project size and complexity and other considerations. Thoughtful changes to program design to align policies and procedures with best practices are recommended.

Based upon gross impact results, determining the appropriate baseline was found to be an important factor driving the discrepancies between ex-ante claimed savings and ex-post evaluated savings for all IOUs. The LRA process also confirmed that baseline determination is a significant area for improvement, particularly for SDG&E. Another important discrepancy factor was operating conditions; the use of pre-installation and post-installation M&V (along with data collection and appropriate adjustments to claimed savings) should reduce the effect of this discrepancy factor. Areas for improvement in pre-installation and post-installation M&V were greatest for SDG&E, SCG, and SCE. Calculation methods were also areas for improvement for these three utilities, and for PG&E to a lesser extent.

IOU-specific findings are summarized in the following bullets:

- PG&E electric projects can improve considerably by applying *appropriate non-HVAC interactive effects methods* and PG&E gas project *measure eligibility* should be improved. Another issue area with widespread performance problems is *measures that exceed code or industry standard practice*.
- SCE had issues with below average performance in the *appropriate HVAC interactive effects method*. Looking across the SCE program groups, two issue areas which

frequently performed below average are *IOU application documentation complete and accurate* and *all relevant inputs considered*.

- SDG&E program assessments show that problems are widespread. All program groups examined have below-average scores for two or more issues related to the *Project Documentation and Specification* category. This entire category was found to be below average for SDG&D overall. This category is an area that requires focused improvement. Other issue areas that demonstrated below average ratings across most of the program groups examined includes *all relevant inputs considered* and *installation meets program rules*.
- SCG exhibits the need for improvement in the *appropriate calculation methods* category. The single issue area of greatest concern is *measures exceed code or industry standard practice*.

The results summarized below are at a program level and can be considered as guidance for further investigation.

- The PG&E core program groups have above average performance with a relatively high overall assessment score and ranking. *Measure eligibility*, however, is an issue area that requires attention.
- The PG&E Industrial Calculated Incentives (core) program performs well above average but PG&E's other third party programs addressing the industrial sector have more limited success. PG&E's Industrial Calculated Incentives program is flagged as below average only for *IOU tracking data completeness and accuracy*. PG&E's other non-core programs in the industrial sector exhibit significant below average areas; there are issues with *IOU application documentation*, *valid RUL/EUL*, *appropriate impact calculation methods*, *measures exceeding code or industry standard practice*, and *installation that meet program rules*.
- New construction programs perform comparatively well in the lower rigor assessments, with the exception of SCE's Savings by Design program, which appears to be below average in the areas *of post-installation M&V* and *project documentation*. SCE should consider more rigorous use of post installation M&V on new construction projects. (Although not always required nor fully specified by program rules, many projects would benefit by incorporating post-installation M&V). The PG&E and SDG&E new construction programs out-perform other program groups, including the LGP, SGP, third party, and core program groups.

- The statewide institutional partnerships[7] have varying, below-average performance across the issues analyzed, and most exhibit a need for improved *project documentation*. The SCE CCC programs are below average in *appropriate calculation methods*. PG&E UC/CSU scores below average in *calculation methods*. SCE UC/CSU programs need some improvement, particularly in *project documentation* and *measure eligibility*.
- PG&E's RCx programs perform consistently better than the overall program average. While *project documentation, pre- and post- installation M&V* and *calculation methods* received above average scores, concern exists in the areas of *measure eligibility*, *industry standard practice*, and *customer installations meeting program rules*.
- PG&E's LGP programs ("Energy Watch") exhibit the need for improvement in the use of *pre- and post-installation M&V, adhering to standard practice guidelines*, and the *use of appropriate impact calculation method*s.
- SCE LGP programs can improve in *project documentation* and *pre- and post-installation M&V.*
- SCE's other third party programs have above average scores in pre- and post-installation M&V *appropriate impact calculation method* and *customer installation meets program rules.* However, this program group scores below average in, *measure eligibility, tracking data,* and *consideration of all relevant inputs.*
- The SCG Commercial Calculated program is significantly below average on several issues: *project utilized pre- and post-installation M&V, appropriate baseline*, and *valid RUL / EUL approach.* It also performs much worse than average on *appropriate impact calculation methods, all relevant inputs considered*, and *adequate values for all inputs*.
- In addition, the SCG Commercial Calculated program shares with the SCG Industrial Calculated program significantly below average scores for *measures exceeding code or industry standard practice*.
- The SDG&E BID program is ranked near the bottom on the overall assessment score and is below average on most issues. This program has the greatest need for improvement in the use of *pre- and post-installation M&V* and *project documentation* where it scores the lowest compared to other SDG&E core and non-core program groups.

It should be noted that not all findings of relative strength or weakness from the lower rigor assessments are supported by final gross impact efforts and may not be reflected in the gross impact results. The core programs, for instance, did not receive very high GRR scores, despite overall favorable results from the LRA process. The new construction programs did have higher mean gross impact scores (although GRRs were highly dispersed and individual project scores

[7] Partnerships include California Community Colleges (CCC), University of California/California State University (UC/CSU), Department of General Services (DGS), and California Department of Corrections (CDCR).

were not generally within 20 percent of unity). The low LRA scores for SCE's new construction program, in this case, did not lead to lower average GRR scores. Likewise, the low LRA scores for SCE and PG&E SGP programs are in contrast to high average GRR scores for those programs. Additional exploratory efforts are required to establish if a clear and robust connection between the LRA and GRR scores, i.e., between the lower rigor assessments and the more rigorous full gross impact work exists.