Itron
Knowledge to Shape Your Future

Electric | Gas | Water
information collection, analysis and application

Final Report

2013 Custom Impact Evaluation Industrial, Agricultural, and Large Commercial

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

With Assistance from:

Energy and Resource Solutions
Energy Metrics
Michaels Engineering
Katin Engineering Consulting

CALMAC Study ID CPU0107.01

July 17, 2015

Table of Contents

List of Tables

List of Figures

1

Executive Summary

This report presents findings from the impact evaluation of the program year (PY) 2013 California Program Administrator (PA)[1] led energy efficiency programs, focusing on nonresidential custom measures.[2] This custom project impact evaluation is one of multiple California Public Utilities Commission (CPUC) evaluations of the PAs' 2013 efficiency programs, and was conducted under the Industrial, Agricultural and Large Commercial (IALC) Roadmap as part of an overarching contract for PY2013-2014 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 PA 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 Project Practices Assessments (PPA) activity[3] to discern possible changes in ex ante savings development practices. Findings and recommendations to improve program performance are also provided.

Three main evaluation activities support the findings and recommendations in this report: (1) M&V activities for estimating gross impacts for 189 projects, (2) telephone survey data collection supporting net to gross (NTG) estimation for a total of 146 projects, and (3) a total of 189 engineering reviews supporting PPA results.

1.1 Custom Impact Evaluation Portfolio Context and Sample Sizes

The programs included in this custom impact evaluation address a wide range of nonresidential facilities, including commercial, institutional, agricultural and industrial applications. The scope addresses nonresidential custom measures of all types with three exceptions: custom lighting

[1] California energy efficiency program administrators include PG&E, SCE, SCG, SDG&E, Marin Clean Energy, the Bay Area Regional Energy Network (REN), and the Southern California REN. However, this evaluation only addresses programs under the administration of PG&E, SCE, SCG and SDG&E.

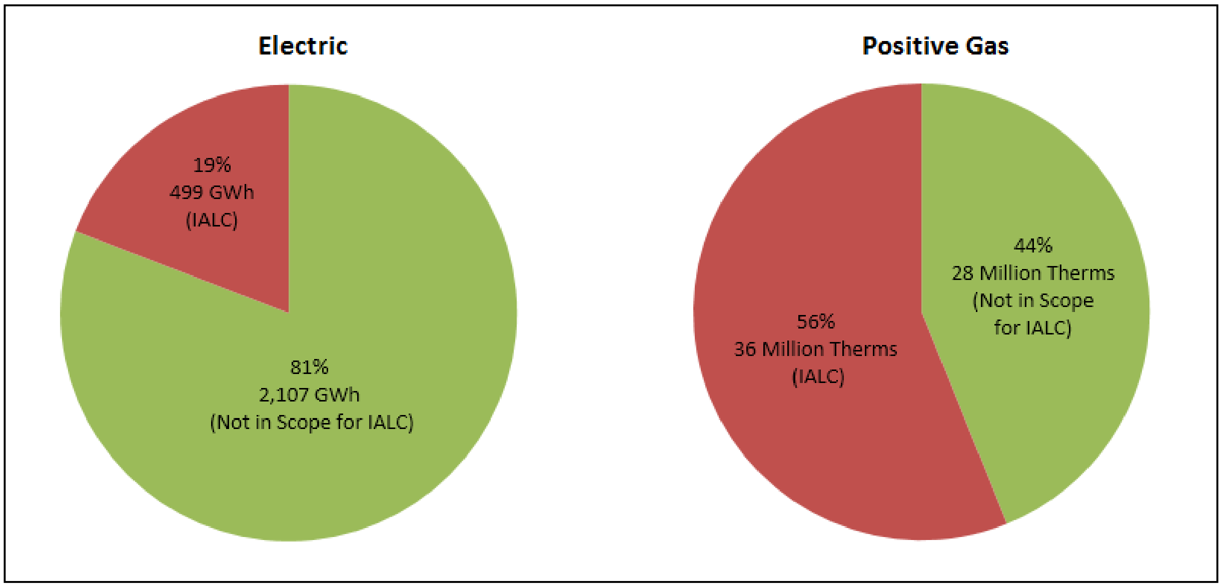
[2] This effort was completed for CPUC under the direction of staff responsible for evaluation of utility energy efficiency programs.

[3] Project Practices Assessment reviews were conducted for all completed M&V points; they feature assessments of project compliance with ex-ante review guidance and requirements, and conformance with policy guidance, with an emphasis on ex-ante gross savings development and methods.

measures, whole building new construction projects and pump test claims.[4] Each custom-oriented PA program offers one or more of the following interventions 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.

As shown in Figure 1-1, energy savings claims associated with the scope of this evaluation represent a significant contribution to the overall savings portfolios for the PAs' energy efficiency programs, accounting for about 19 percent of statewide electric savings claims and 56 percent of statewide gas savings claims[5] during PY2013. During this period, the PA tracking data for measures associated with this custom impact evaluation included thousands of records statewide with annual electric savings claims by the PAs totaling 499 GWh and annual gas savings claims totaling 36 million Therms.

Figure 1-1: Custom Impact Evaluation Share of Statewide PY2013 Energy Efficiency*



* "Positive" gas refers to exclusion, in this chart, 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. Ultimately, due to the number of gross impact M&V and net impact NTG sample points targeted for the study, and the number of sample points required to provide reasonable statistical precision for a sampling

4 Custom lighting measures are addressed in a separate impact study on nonresidential lighting, the CPUC 2013-2014 Commercial Roadmap. Whole building new construction projects are addressed under a separate work order falling under the IALC Roadmap. Pump test claims were evaluated in the past and are not in scope for 2013.

5 Excluding negative claims associated with HVAC interactive effects.

domain, the primary sampling domains for developing and reporting gross and net impact results were each of the PA territories on a combined MMBtu basis, where applicable. This approach resulted in the following four sampling domains for which gross realization rates and NTG ratios were developed and reported: PG&E (electric and gas combined,) SCE (electric and gas combined,) SDG&E (electric and gas combined,) and SCG (gas only). The custom evaluation collected information from 189 gross impact points (consisting of 240 individual measures) and 146 net impact points, where a point or "sampling unit" is defined as an individual project (application) installed at a specific site. The original sample design was to have 153 NTG projects fully nested within a 190 point M&V sample. However, given customer willingness to participate and other factors, the final gross and net samples did not fully align. In total, 111 of 146 NTG points were fully nested within the gross M&V sample. The total sample size (including main and backup points achieved) and percent of ex-ante MMBtu claims by PA are shown in Table 1-1 below.

Table 1-1: Summary of Custom Evaluation Sample Sizes by PA

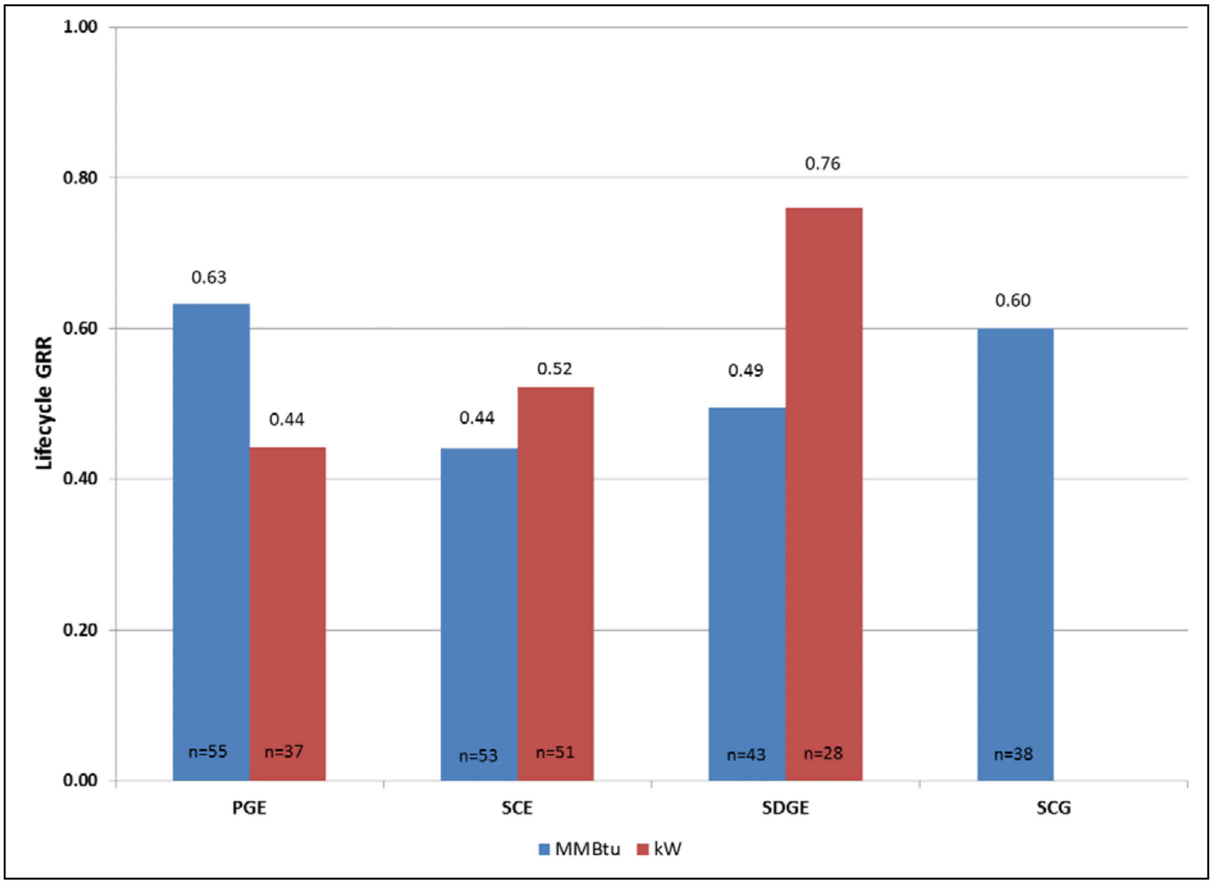
PA	Completed M&V Points (n)		Completed NTG Points (n)		Percent of Ex-Ante MMBtu Claims	
	Main	Backup	Main	Backup	M&V Sample	NTG Sample
PG&E	53	2	34	17	33%	33%
SCE	52	1	30	16	31%	29%
SDG&E	41	2	22	6	48%	41%
SCG	35	3	17	4	76%	60%
All PAs	**189**		**146**		**41%**	**37%**

1.2 High-Level Custom Gross Impact Results

Figure 1-2 and Table 1-2 below summarize the mean lifecycle gross impact realization rates (GRRs) for each of the four PA sample domains. Gross realization rates are calculated for each sampled project as the ex-post, evaluation based estimate of savings divided by the PAs' ex-ante estimate of savings. 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.

The mean lifecycle realization rates by PA and energy metric are less than 0.70 for all but one energy metric (the GRR for SDG&E kW is 0.76) and are similar to those from 2010-2012 (see weighted MMBtu comparison in Chapter 3). All weighted lifecycle realization rates are lower than the corresponding first year realization rate.[6] Generally, evaluation lifecycle realization rates remain significantly below the 0.9 default ex-ante GRR adjustment for custom programs. A significant number of projects (31 out of 189) were estimated to have negative and/or zero GRRs.

Figure 1-2: Mean Lifecycle Gross Realization Rates by PA and Energy Metric (MMBtu and kW)



The error ratios for most domains (Table 1-2) are similar to the error ratios obtained in the *2010-2012 WO033 Custom Impact Evaluation Final Report.*[7] In the 2013 sample, the relatively high standard deviation (e.g. MMBtu GRR for SCE) is indicative of the variability in the data rather than of a small sample size. The underlying sample has individual gross realization rates that are widely dispersed between zero and values exceeding 1.0, which will always result in a large standard deviation regardless of the number of projects sampled. While the precision of the 2013 results are similar to previous evaluation results, the reader should be cognizant of the relatively broad confidence intervals when interpreting the results and findings.

6 Lifecyle gross realization rate results are pushed substantially downwards relative to first year results by differences between evaluation and ex-ante EUL specification.

7 http://www.calmac.org/publications/2010-12_WO033_Custom_Impact_Eval_Report_Final.pdf

Table 1-2: Mean Lifecycle Gross Realization Rates by PA and Energy Metric (MMBtu and kW)

Energy Metric	Sample Size (n)	Mean LC Gross Realization Rate	Population (N)	Error Ratio**	90% Confidence Interval	FY Mean GRR
PG&E						
MMBtu*	55	0.63	1,125	0.47	0.57 to 0.70	0.74
kW	37	0.44	853	1.41	0.28 to 0.61	0.54
SCE						
MMBtu*	53	0.44	934	1.07	0.34 to 0.54	0.54
kW	51	0.52	838	0.80	0.43 to 0.62	0.64
SDGE						
MMBtu*	43	0.49	264	0.88	0.40 to 0.59	0.75
kW	28	0.76	144	0.90	0.57 to 0.95	1.02
SCG						
MMBtu*	38	0.60	158	0.86	0.48 to 0.72	0.69

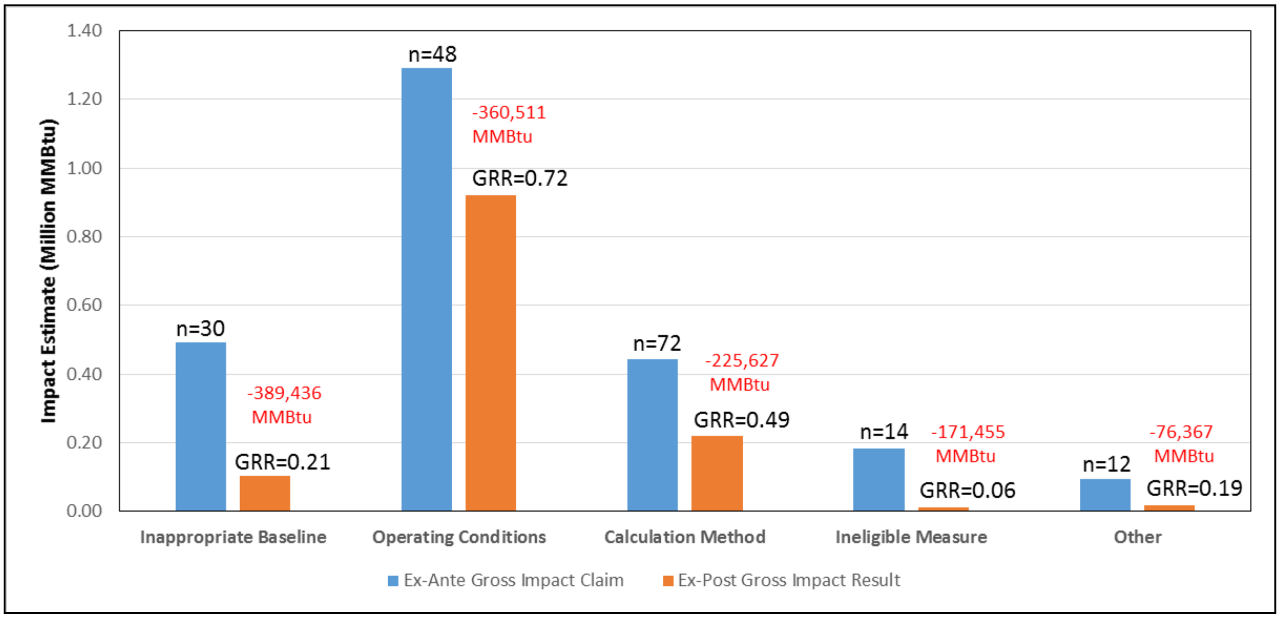
* The primary sample was designed and selected at this level. The kW sample sizes are lower due to the fact that kW impacts were not claimed by PAs in every case.

** 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) baseline specifications, (2) observed operating conditions, and (3) the PA's calculation methods. These discrepancy factors were examined for all projects where they caused upward or downward adjustments to the ex-ante savings. Of the 240 records (measures) studied, these discrepancy factors explain a portion of the downward adjustments in ex-ante savings for 13 percent, 20 percent and 30 percent of records, respectively.[8] For all downward adjustments, across all PAs, these three factors reduced their respective ex-ante MMBtu savings by 79 percent (baseline specification), 51 percent (calculation methods), and 28 percent (operating conditions) (Figure 1-3).

8 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.

Figure 1-3: Summary of Discrepancy Factors Resulting in Downward Adjustments to Ex-Ante MMBtu Impacts - All PAs



1.3 High Level Custom Net-to-Gross Results

NTGR results at the PA level are presented in Figure 1-4 and Table 1-3.

Figure 1-4: Weighted Net-to-Gross Ratios by PA[9]

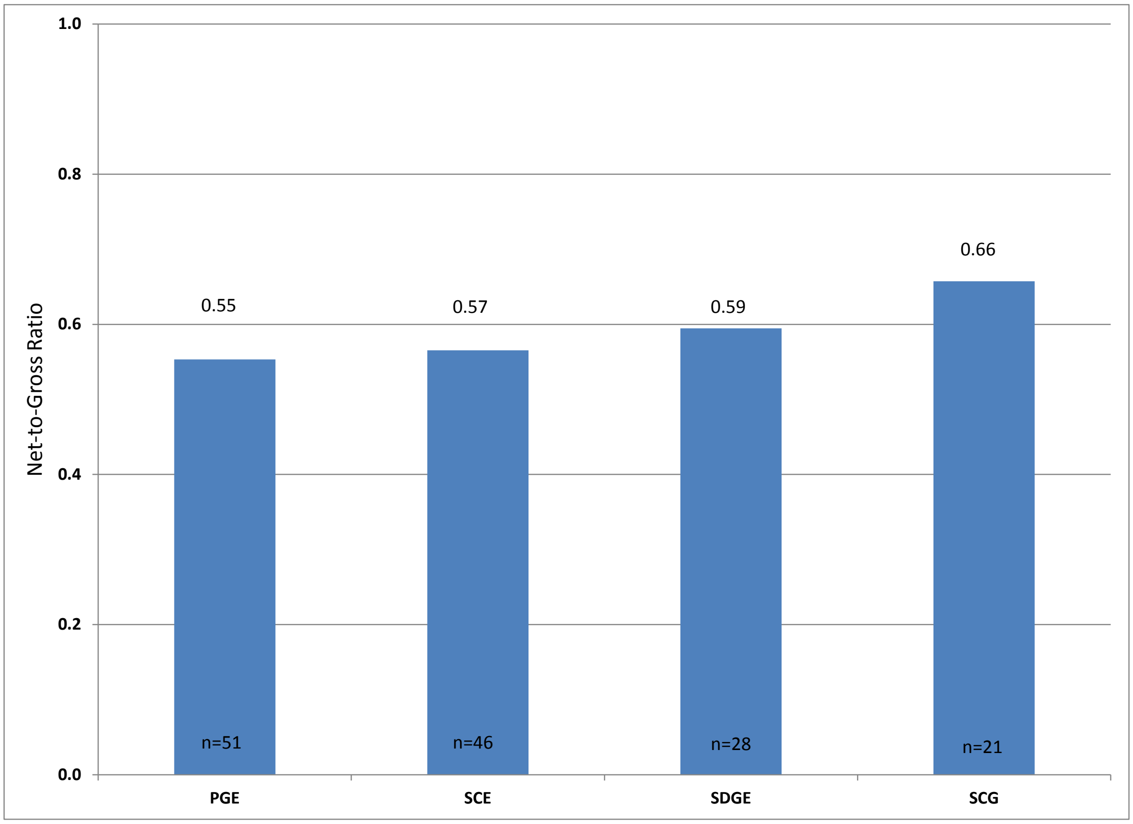


Table 1-3: Weighted Net-to-Gross Ratios by PA

Results	Mean Net-to-Gross Ratios			
	PGE	SCE	SDG&E	SCG
Weighted NTGR	**0.55**	**0.57**	**0.59**	**0.66**
90 Percent Confidence Interval	0.52 to 0.59	0.52 to 0.61	0.55 to 0.64	0.59 to 0.73
Relative Precision	0.07	0.09	0.08	0.10
n NTGR Completes	51	46	28	21
N Sampling Units	1,126	934	264	159
Error ratio (ER)	0.30	0.36	0.28	0.31

9 Note that these values reflect the removal of 15 projects from NTG calculations. 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 (13 projects removed) or Dual Baseline (2 projects removed) determinations in the gross impact analysis. For all PA domains, the change in domain level NTGRs from these removals was very slight, on the order of one to two percent.

Based on the NTGR results presented above, the following observations stand out:

- At the level of PA sampling domain, the final NTGRs range from 0.55 to 0.66, equating to about a 20 percent improvement over the previous cycle (see weighted MMBtu comparison in Chapter 4).
- **PG&E:** The weighted NTGRs for the PG&E improved by 8 percent compared to PY2010-2012 evaluation results.
- **SCE:** Current cycle results have improved by 16 percent compared to PY2010-2012, based on an NTGR of 0.57 for PY2013 projects versus an NTGR of 0.49 for PY2010-2012.
- **SDG&E:** NTGRs for SDG&E's projects have improved by 23 percent compared to PY2010-2012 results, averaging 0.59 across all PY2013 projects evaluated.
- **SCG:** For SCG the weighted NTGR across all projects is 0.66. This represents a 35 percent improvement over the PY2010-2012 average NTGR of 0.49.

Behind 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 by PA. 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.

As in PY2010-2012, 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. In addition to corporate policy, certain factors can contribute to lowering the level of program influence at a project level (e.g. a project that is in the capital and/or operating budget; a measure that is already installed elsewhere in the company, in places that do not offer rebates; a measure that is associated with environmental compliance, etc.). Other factors also can contribute to increasing program influence at the project level (e.g. timing of decision relative to discussions with program staff; first-time installation of a measure by the company; program incentive is a high percentage of the project cost, etc.).

1.4 Net Evaluation Realization Rate Results

Net evaluation realization rates are presented for each PA in Table 1-4 through Table 1-7. Net realization rates are the product of the GRRs and the NTGRs, and thus portray the combined evaluation impact as compared to unadjusted PA ex-ante gross impact claims. These tables also provide a comparison of the ex-post net realization divided by the PAs' net realization rates (that is, the evaluation results compared to the PAs' ex-ante values inclusive of the default GRR of 0.9 and the PAs' ex-ante NTG values).

Please note that all projects that have been subject to ex-ante review (EAR), and that are subsequently installed, can be fully claimed by the PAs (in other words: PA RR=1.0). To claim all other non-deemed projects, PAs adjust ex-ante estimates by RR=0.9 as ordered by the CPUC in D.11.07.030. A total of 56 EAR projects were part of the IALC 2013 population: 38 were installed in PG&E territory, 12 in SCE territory, four in SDG&E territory, and two in SCG territory. This explains why the claimed GRR from line b. in the tables that follow is higher than 0.90 in some cases. Seven of these 56 projects were randomly sampled and were analyzed by the IALC Custom evaluation (refer to Chapter 3 for results).

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

Impact Element	LC Electric Savings		LC Gas Savings
	kWh	Avg. Peak kW	Therms
Tracking			
a. Claimed LC Gross Savings	3,070,569,225	412,726	293,168,434
b. Claimed GRR	0.91	0.90	0.90
c. Claimed Adjusted Gross Savings (c = a x b)	2,784,196,129	373,014	264,522,775
d. Claimed NTGR	0.65	0.64	0.62
e. Claimed Net Savings (e = c x d)	1,812,308,334	240,015	165,122,455
f. Claimed Net Realization Rate (f = b x d)	0.59	0.58	0.56
Evaluation			
g. Evaluation LC GRR	0.63	0.44	0.63
h. Evaluated Gross Results (h = a x g)	1,944,584,951	182,310	185,662,945
i. Evaluation NTG Ratio	0.55	0.55	0.55
j. Evaluated Net Results (k = h x i)	1,075,748,854	100,854	102,709,167
k. Evaluation Net Realization Rate (l = g x i)	0.35	0.24	0.35
l. Evaluated Net Savings as a Fraction of Claimed Net Savings (m = k / f)	0.59	0.42	0.62

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

Impact Element	LC Electric Savings		LC Gas Savings
	kWh	Avg. Peak kW	Therms
Tracking			
a. Claimed LC Gross Savings	2,605,321,450	363,824	3,707,155
b. Claimed GRR	0.90	0.90	0.90
c. Claimed Adjusted Gross Savings (c = a x b)	2,352,054,124	328,756	3,336,479
d. Claimed NTGR	0.60	0.60	0.80
e. Claimed Net Savings (e = c x d)	1,409,425,957	198,586	2,673,805
f. Claimed Net Realization Rate (f = b x d)	0.54	0.55	0.72
Evaluation			
g. Evaluation LC GRR	0.44	0.52	0.44
h. Evaluated Gross Results (h = a x g)	1,148,010,870	189,850	1,633,524
i. Evaluation NTG Ratio	0.57	0.57	0.57
j. Evaluated Net Results (k = h x i)	648,952,122	107,319	923,405
k. Evaluation Net Realization Rate (l = g x i)	0.25	0.29	0.25
l. Evaluated Net Savings as a Fraction of Claimed Net Savings (m = k / f)	0.46	0.54	0.35

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

Impact Element	LC Electric Savings		LC Gas Savings
	kWh	Avg. Peak kW	Therms
Tracking			
a. Claimed LC Gross Savings	648,105,655	51,553	13,711,377
b. Claimed GRR	0.90	0.90	0.90
c. Claimed Adjusted Gross Savings (c = a x b)	585,487,365	46,556	12,342,439
d. Claimed NTGR	0.61	0.60	0.64
e. Claimed Net Savings (e = c x d)	354,352,016	28,143	7,859,988
f. Claimed Net Realization Rate (f = b x d)	0.55	0.55	0.57
Evaluation			
g. Evaluation LC GRR	0.49	0.76	0.49
h. Evaluated Gross Results (h = a x g)	320,532,336	39,213	6,781,209
i. Evaluation NTG Ratio	0.59	0.59	0.59
j. Evaluated Net Results (k = h x i)	190,558,641	23,312	4,031,474
k. Evaluation Net Realization Rate (l = g x i)	0.29	0.45	0.29
l. Evaluated Net Savings as a Fraction of Claimed Net Savings (m = k / f)	0.54	0.83	0.51

Table 1-7: SCG Lifecycle Net Realization Rate Estimates and Comparisons

Impact Element	LC Gas Savings
	Therms/year
Tracking	
a. Claimed LC Gross Savings	191,615,864
b. Claimed GRR	0.91
c. Claimed Adjusted Gross Savings (c = a x b)	174,381,245
d. Claimed NTGR	0.50
e. Claimed Net Savings (e = c x d)	87,865,120
f. Claimed Net Realization Rate (f = b x d)	0.46
Evaluation	
g. Evaluation LC GRR	0.60
h. Evaluated Gross Results (h = a x g)	115,078,042
i. Evaluation NTG Ratio	0.66
j. Evaluated Net Results (k = h x i)	75,651,044
k. Evaluation Net Realization Rate (l = g x i)	0.39
l. Evaluated Net Savings as a Fraction of Claimed Net Savings (m = k / f)	0.86

1.5 Summary of Findings and Recommendations

This report provides findings and recommendations aimed at improving custom program performance and supporting PA program design and procedure enhancements for this important element of the PAs' energy efficiency portfolios. Findings and recommendations were developed from each of the primary analysis activities: impact evaluation, net evaluation, Program Practices Assessment (PPA) activities. Extensive reporting of findings and recommendations is presented in Chapter 6 of this report.

At a summary level, the detailed recommendations in this report fall into the following primary areas:

- To more accurately estimate ex-ante savings, the PAs should:
  - Improve documentation and reporting of project EUL,[10] including a review of evaluation EUL conclusions/rationale in an effort to improve EUL claims and LC GRR results,

[10] It is notable that the evaluation estimate of EUL differed from the PAs estimate 46 percent of the time. For those instances the evaluation-derived average EUL was smaller than the ex-ante average EUL by more than four years, representing a 35 percent reduction in the ex-ante EUL claim for that subset of observations. As noted in Chapter 3 LC GRR results were substantially lower than FY GRR results and this EUL difference was *the key factor* driving down the LC GRR results. As described in Chapter 5, primary drivers of EUL differences include different

  - Improve quality control of project operating conditions, ex-ante baseline determinations, savings calculations, and eligibility rules to address the discrepancy factors presented in this report, and
  - Ensure adjustments to project savings based on post-installation inspections and M&V.
- To achieve sufficient quality control, PAs should increase due diligence on accuracy, comprehensiveness and documentation in project application files. SDG&E appears to be leading the other PAs in this area, given the level of improvement noted in Chapter 5 in 2013+ applications.
- To reduce continued moderate free ridership, PAs should test changes to program features designed to increase program-induced savings.

Finally, key recommendations discussed in Chapter 6 of this report are listed in Table 1-8.

Table 1-8: 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
The evaluation team recommends that the PAs ensure that savings calculations are based on actual equipment-use schedules and reflect any changes to the post-installation operating parameters (such as flow rates, temperatures and set points, system pressures, production rates, power measurements)
Baseline Conditions
Increase efforts to ensure conformance with CPUC baseline policies and make a greater effort to examine existing equipment RUL
Clearly identify project event in terms of natural replacement, replace on burnout, early replacement, new construction, and add-on equipment, and set the appropriate baseline accordingly
Disseminate information on baseline selection to ensure best practices across program staff, implementers and customers

interpretations of applications of DEER, EUL limited by RUL of host equipment for add-on / systems optimizations, and RCx program rules.

Key Recommendations by Topic Area
The PAs need to do a better job of ensuring that baseline equipment specifications are capable of meeting post-installation operating requirements, that the baseline selected is consistent with the project type, and that regressive baseline considerations are examined
The requirement that measures exceed the ISP / code baseline is a first order consideration for project eligibility
Calculation Methods
Continue to review and improve impact methods and models through review of evaluation results, industry best practices, and the CPUC's ex-ante review process
Ex-ante savings estimates and calculation methods should be more thoroughly reviewed and approved by PA technical staff prior to finalization of incentives and savings claims
Conduct periodic due diligence to ensure programs adhere to PA and CPUC impact estimation policies, guidelines, and best practices
Continue to work closely and collaboratively with the CPUC's ex-ante review process
Cross-Cutting and Other Gross Impact-Related
Improve PA program requirements, manuals, training, and quality control procedures in order to screen out ineligible projects
It is recommended that the PAs carefully review each of the 30 Final Site Reports (FSRs) listed in Table 3-4 to identify the specific reasons that led zero or negative savings, and use those lessons learned to improve related project practices
It is recommended that a statewide document, similar to the PPA form, be developed for use by all PAs for custom claims; this form has been provided to the PAs for their use
The PA's project eligibility treatment suggests that the PA's communication and coordination efforts with entities responsible for disseminating and implementing CPUC guidance should be increased
Better ex-ante documentation is needed supporting key project parameters identified in the PPA chapter
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

Key Recommendations by Topic Area
Make changes to the incentive design by setting incentive levels to maximize net (not gross) program impacts
Use a comprehensive mix of program features and leverage an array of delivery channels to engage individual customers and encourage a long-term energy efficiency-based focus
More information should be developed 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 draft results from the impact evaluation of the 2013-2014 California Program Administrator's (PAs)[11] energy efficiency programs, focusing on nonresidential custom measures. This effort is managed by the CPUC Energy Division (ED) staff and is referenced as Itron Industrial, Agricultural and Large Commercial (IALC) Roadmap on the CPUC ED public documents website.[12] The *IALC Research Plan* dated July 1, 2014, provides additional detail on the evaluation effort; this evaluation plan is available on the ED public documents website.[13] Readers may also want to familiarize themselves with a number of other relevant CPUC sources that are referenced throughout this report.[14] This includes a nonresidential Net-to-Gross (NTG) methods document,[15] which can also be found on the ED public documents website.[16] The scope of work includes independent estimation of gross and net savings, and development of findings and recommendations that can be used to improve program and measure effectiveness.

This chapter provides background information and introduces the reader to the types of programs, facilities, and interventions evaluated under the IALC roadmap. This chapter also references the research plan and evaluation procedures at a very high level: additional study background is provided, highlighting the percentage of portfolio claimed savings associated with the IALC evaluation effort, and presenting the study objectives and issues researched.

---

[11] California energy efficiency program administrators include PG&E, SCE, SCG, SDG&E, Marin Clean Energy, the Bay Area Regional Energy Network (REN), and the Southern California REN. However, this evaluation only addresses programs under the administration of PG&E, SCE, SCG and SDG&E.

[12] http://www.energydataweb.com/cpuc/home.aspx

[13] http://www.energydataweb.com/cpuc/search.aspx, "Custom Impact Evaluation – 2013-2014 – DRAFT RESEARCH PLAN

[14] It should be noted that this evaluation report is results-focused, referring readers to other supporting documents and appendices to further address methods, CPUC guidelines, supporting studies and procedures. References to supporting documents and appendices generally appear at the front of each chapter.

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

[16] The NTG methods document was distributed and discussed with PA project coordination group (PCG) and evaluation staff during previous evaluation efforts, 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. In addition, Project Practices Assessments (PPAs) examine custom project impact estimation methods and procedures, and facilitate an assessment of PA ex-ante performance for custom projects. These reviews feature assessments of project compliance with ex-ante review guidance and requirements, and conformance with policy guidance, with an emphasis on ex-ante gross savings development and methods.

More than 100 of the PY2013-2014 utility programs[17] include non-residential, non-deemed (custom) projects. Some programs, such as the PA 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 PA programs, including those undertaken by third parties or through local government partnerships, with the main objective to estimate PA-specific realization rates for custom projects across programs.

A goal of this impact evaluation is to provide the PAs with feedback that can be used to make any necessary corrections to improve their current programs, as well as feedback on what aspects of program design and implementation are successful. For this to be successful, evaluations must be timely in their reporting of results and feedback to allow utility staff adequate time for making necessary improvements. This IALC impact report addresses findings for the 2013 claim period; another IALC report will be prepared once the 2014 claim period is evaluated.

The CPUC organized all of its consultant evaluation and research work for PY2013-2014 into roadmaps. Some of these roadmaps address specific measures, sectors, or programs, while others address 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 PA’s portfolio were mapped to a measure group. Measure groups were then mapped and assigned to different roadmaps, each of which has its own project team, scope, and reporting. Mapping of assignments to road maps was also informed by residential versus nonresidential participation, deemed versus non-deemed (i.e., custom), upstream versus downstream provision of incentives, and other considerations. The IALC roadmap was assigned all of the nonresidential custom projects, excluding lighting and codes and standards claims.

[17] In 2013 PG&E had 51 programs that include custom projects, SCE had 42 such programs, SDG&E had 11, and SCG had 4.

Energy savings claims from the measures assigned to the IALC roadmap represent a significant contribution to the overall savings portfolios for the PA's 2013 energy efficiency programs, accounting for about 19 percent of statewide electric savings claims and 56 percent of statewide positive gas savings claims. In 2013 the PA tracking data for measures assigned to the IALC roadmap included thousands of entries statewide with annual electric savings claims by the PAs totaling nearly 500 GWh and nearly 77 MW. Statewide PA positive gas savings claims for measures assigned to the IALC roadmap total 36 million Therms.

The most recent PA data extract, which reflects cumulative PA program activity through Q3, 2014, was used to summarize the 2013 claimed energy savings associated with the PA portfolios, as well as the savings assigned to the IALC roadmap's custom impact.[18] These savings are reported in Table 2-1.

Table 2-1: 2013 Claimed Energy Impacts for Projects in the IALC Roadmap, and in the Portfolio, by PA

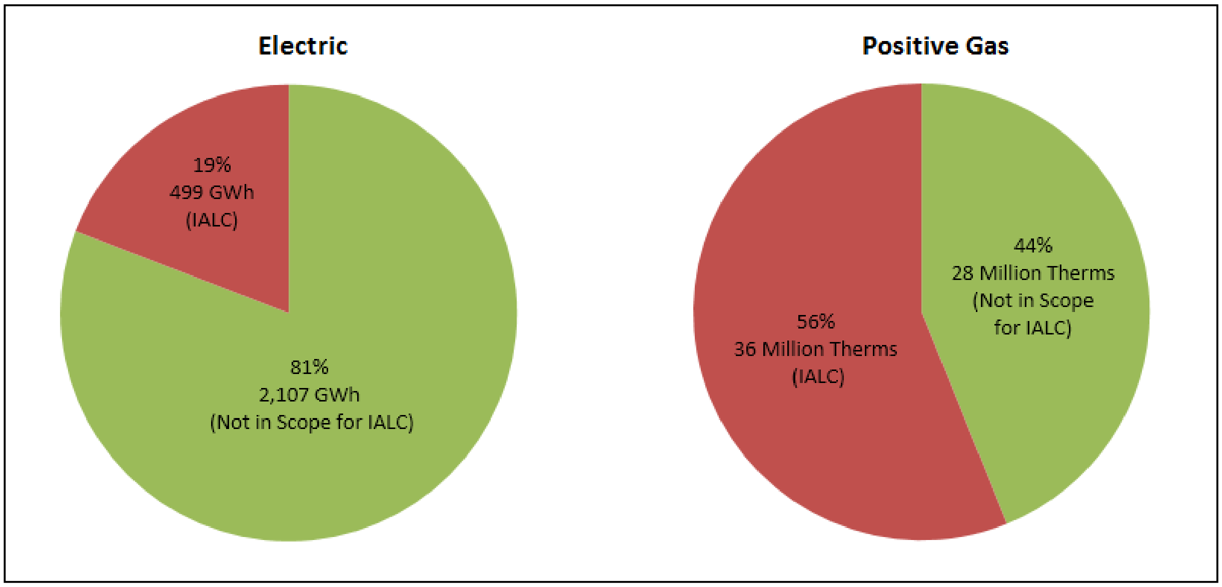
Claimed Impacts by PA			
PA	**Electric Energy (GWh)**	**Electric Demand (MW)**	**Gas Energy (Million Therms)**[19]
2013 PA Savings Claims, IALC Roadmap			
PG&E	235	37	21
SCE	215	34	1
SCG	-	-	13
SDG&E	50	6	1
Total	499	77	36
Total 2013 PA Savings Claims			
PG&E	1,490	242	35
SCE	786	142	1
SCG	8	6	26
SDG&E	322	47	2
Total	2,606	436	64
IALC Percentage of Total PA Savings Claims			
PG&E	16%	15%	60%
SCE	27%	24%	56%
SCG	0%	0%	52%
SDG&E	15%	12%	51%
Total	19%	18%	56%

[18] 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.

[19] Gas savings reported includes only tracking records with positive Therm impacts. A significant number of negative records in the PA portfolio are associated with increased heating due to the interactive effects of lighting efficiency measures. The IALC evaluation addresses only positive Therm saving claims. Negative Therm records were therefore not included in the Table 2-1 energy saving claims summary.

Figure 2-1 contrasts the IALC roadmap savings claims with total portfolio claims for 2013.

Figure 2-1: 2013 IALC Roadmap Savings Claims Relative to Portfolio Claims



The IALC evaluation is further divided into two Work Orders in an effort to isolate and report separately on whole building new construction projects:

- The Custom Impact Evaluation — featuring appropriately rigorous M&V and NTG evaluation — addresses gross and net impact measurements after isolating a heterogeneous subset of non-deemed claims in the IALC population.
- The Nonresidential New Construction Whole Building (NRNC) Evaluation features M&V and NTG evaluation similar to the Custom Impact Evaluation, but applied to whole-building new construction projects.

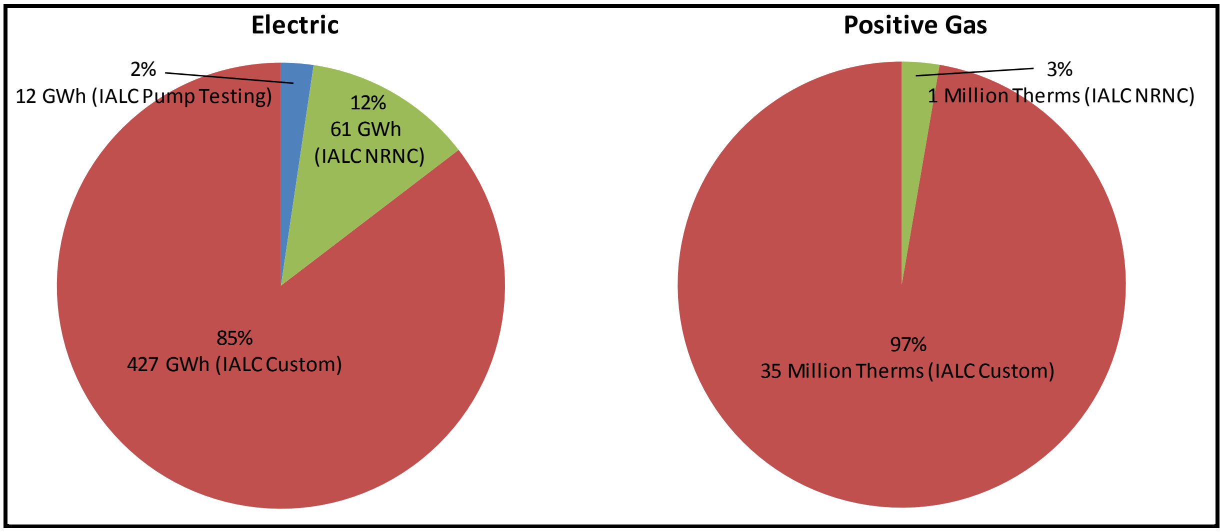
During the process of grouping the IALC claims into Custom and NRNC, a separate group of projects emerged. Through its Agriculture Energy Advisor Program (SCE-13-SW-004A), SCE customers can benefit from full service pump efficiency improvement services (a.k.a. agricultural pump testing). This measure was evaluated in the 2006-2008 program cycle and is not currently in scope for the 2013 evaluation. The 4,986 records corresponding to SCE pump testing that were assigned to IALC roadmap in 2013 were grouped separately from the Custom evaluation. Table 2-2 shows how the records assigned to IALC roadmap in 2013 were further separated into the NRNC, Pump Testing, and Custom groups. Only positive gas claims are shown.

Table 2-2: 2013 Claimed Energy Impacts for the IALC Roadmap, by Group and PA

Claimed Impacts by PA			
PA	**Electric Energy (GWh)**	**Electric Demand (MW)**	**Positive Gas Energy (Million Therms)**
IALC 2013 PA Savings Claims			
PG&E	235	37	21
SCE	215	34	1
SCG	0	0	13
SDG&E	50	6	1
Total	499	77	36
IALC 2013 Whole Building NRNC PA Savings Claims			
PG&E	32	10	1
SCE	20	6	0
SCG	0	0	0
SDG&E	9	3	0
Total	61	20	1
IALC 2013 Pump Testing PA Savings Claims			
PG&E	0	0	0
SCE	12	3	0
SCG	0	0	0
SDG&E	0	0	0
Total	12	3	0
IALC 2013 Custom PA Savings Claims			
PG&E	203	27	20
SCE	183	25	0
SCG	0	0	13
SDG&E	40	3	1
Total	427	55	35

Figure 2-2 presents the Custom portion of the IALC roadmap ex-ante savings claims that were included in the 2013 evaluation and that will be presented in the remainder of this report. The figure also shows the NRNC portion that was evaluated and reported elsewhere, as well as the Pump Testing portion that is not in scope in 2013. Only positive gas claims are shown.

Figure 2-2: 2013 Saving Claims Assigned to the IALC Roadmap



Please note that Table 2-1, Table 2-2, Figure 2-1, and Figure 2-2 reflect PA *claimed* gross savings, i.e. PA savings claims adjusted by the 90 percent default PA RR for all records not affected by the EAR process.[20] To facilitate a fair context for, and comparison point to, Custom evaluation results, all Custom savings claims shown in the remainder of the report and appendices reflect gross (unadjusted) savings claims. That is, unless otherwise noted, the remainder of the report refers to ex-ante savings that are calculated as PA savings claims from which the 90 percent default PA RR is backed out. In addition, as specified in the *IALC Research Plan*, only positive unadjusted savings claims were evaluated (refer to Table 2-3 for totals).

Table 2-3: 2013 Unadjusted Positive Gross Savings Claims Evaluated by IALC Custom, by PA

PA	Positive Electric Energy (GWh)	Positive Electric Demand (MW)	Positive Gas Energy (Million Therms)
IALC 2013 Evaluated Unadjusted Positive Gross Savings			
PG&E	224	29	22
SCE	202	28	0
SCG	0	0	14
SDG&E	45	3	1
Total	471	61	38

[20] The EAR process involves an M&V-level of review for PA 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 PA within-program engineering processes and procedures more generally. Projects that are subject to the EAR process and are subsequently installed and claimed by the PAs are not subject to further adjustment by the 90 percent default RR (in other words: PA RR=1.0 for these projects.

2.2 Study Objectives and Researchable Issues

The overarching goals and objectives for the IALC Custom evaluation are: to verify and validate the energy efficiency savings claims reported from PA 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) were estimated and compared to PA savings claims using evaluation-based realization rates and NTG ratios.

The priorities for this evaluation effort and the researchable issues that this evaluation seeks to examine are described as follows:

1. Estimating the level of achieved gross impact savings, determining what factors characterize gross realization rates, and, as necessary, assessing how realization rates can be improved.
2. Estimating the level of free ridership, determining the factors that characterize free ridership, and, as necessary, providing recommendations on how free ridership might be reduced.[21]
3. Providing timely feedback to PAs to facilitate program design improvements.
4. Determining whether the impact estimation methods, inputs, and procedures used by the PAs and implementers are consistent with the CPUC's policies, decisions and best practices.[22] The evaluation identifies issues with respect to impact methods, inputs, and procedures and makes recommendations to improve PA savings estimates and realization rates.
5. Improving baseline specification, including collecting and reporting on dual baseline. Estimating the extent of any program-induced acceleration of replacement of existing equipment and, in such cases, the RUL of the pre-existing equipment.
6. Collecting data and information to assist with other research or study areas, which could include measure cost estimation, cost effectiveness, updates to DEER, strategic planning, and future program planning.

[21] The IALC Custom NTG surveys also include a battery of questions to address participant spillover. However, these data are analyzed and reported on as part of the 2013-14 Nonresidential Spillover Study under the Residential Roadmap and Market Studies PCG. The 2013-14 Nonresidential Spillover Study evaluation plan can be found at: http://www.energydataweb.com/cpucFiles/pdaDocs/1235/PY2013-2014%20Non-Res%20SO%20Evaluation%20Plan%202015_02_10.pdf

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

In order to more fully answer these researchable questions, the Custom evaluation collected information from 189 gross impact points and 146 net impact points, where a point or "sampling unit" is defined as an individual project (application) installed at a specific site. Gross impact estimates for sampled projects were based on field inspections, measurements, and extensive engineering analysis (i.e., M&V); the gross impact results are discussed in Chapter 3. The NTG evaluation consisted of interview-based evaluation of a representative sample of selected projects, and the use of ratio estimation to aggregate to domain-level net savings estimates; the net impact results are discussed in Chapter 4.

In addition, Project Practices Assessments (PPAs) were incorporated into the 2013-2014 IALC Custom Impact analysis. The purpose of the PPA process is to build upon the results of the Low Rigor Assessment (LRA) process that was part of the 2010-2012 evaluation. To examine the influence of the EAR process on program results, the PPA process was based on all sampled gross impact points, but analyses and feedback were conducted separately for applications with a customer agreement date falling in 2013 (i.e. those projects expected to be influenced by EAR recommendations) versus all earlier applications.[23] The results of the PPA analysis are discussed in Chapter 5.

Given the expected range of error ratios (coefficient of variation for a ratio estimator) for the gross realization rates (roughly 0.6 to 1.0 based on the 2010-2012 custom impact evaluation), and the small number of impact (M&V) and NTG points implemented, only a relatively small number of sampling domains could be supported for the 2013 study. Since the IALC Custom evaluation was expected to provide results at PA-level, M&V and NTG samples were designed and implemented at PA-level.

To allow evaluation of both electric and gas projects in a single domain (each) for PG&E and SDG&E, kWh electric savings and Therms gas savings at the project level were converted into source energy (MMBtu) savings for stratification and sampling purposes.[24] Sampling and analysis on the basis of source energy savings were conducted for SCE and SCG as well, for consistency in reporting and easy comparison of results across the PAs.

Analysis of M&V and NTG data yields weighted MMBtu gross realization rates (GRRs) and net-to-gross ratios (NTGRs) for each PA, as well as weighted kW GRRs for PG&E, SCE and SDG&E. The MMBtu GRRs and NTGRs were used to estimate both electric kWh ex-post savings and gas

[23] Project Practices Assessment reviews feature assessments of project compliance with ex-ante review guidance and requirements, and conformance with policy guidance, with an emphasis on ex-ante gross savings development and methods.

[24] Conversion rates obtained from "*2001 Energy Efficiency Standards for Residential and Non-residential Buildings, California Energy Commission*," June 2001: 1 kWh = 10,239 Btu source energy; 1 Therm = 100,000 Btu source energy. 1 MMBtu =1,000,000 Btu.

Therm ex-post savings for each PA. Ex-post kW savings for each PA were estimated by using kW GRRs.

2.3 Structure of the Report

Table 2-4 shows the overall organizational structure of this report. Although findings and recommendations are overarching in Chapter 6, it is noteworthy that findings are also highlighted in Chapters 3, 4 and 5. Readers seeking a more comprehensive assessment of opportunities for program improvement, and the details and reasons behind findings, are therefore encouraged to read these particular chapters.

Table 2-4: 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	**Goss Impact Results**	Gross impacts and realization rates, causes and effects of ex-ante and ex-post impact differences, and ex-post suggestions and considerations for ex-ante estimation improvement
4	**Net Impact Results**	Net of free ridership ratios and results, and key factors influencing NTGRs
5	**PPA Results**	Assessments based on a comparison between ex-ante and ex-post M&V-based conclusions and finding
6	**Detailed Findings and Recommendations**	Presented by Chapter 3-5 topic areas and appropriate sub-topics that are examined in those chapters

3

Gross Impact Results

This chapter presents quantitative and qualitative gross impact results for the 2013 IALC 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, Program Administrator (PA)[25] domain and size stratification. Results are also presented for energy metrics – source energy (MMBtu)[26] and electric demand (kW).

Stratified sampling was implemented separately for custom measures installed in 2013 by each PA: PG&E, SCE, SDG&E and SCG (for more detail please refer to the Custom Evaluation Research Plan referenced in Chapter 2). Unless noted otherwise, gross realization rates represent the full lifecycle of the projects examined, that is, the lifecycle ex-post evaluation-based estimate of impacts divided by the PAs' lifecycle ex-ante estimate of impacts.[27]

---

[25] California energy efficiency program administrators include PG&E, SCE, SCG, SDG&E, Marin Clean Energy, the Bay Area Regional Energy Network (REN), and the Southern California REN. However, this evaluation only addresses programs under the administration of PG&E, SCE, SCG and SDG&E.

[26] Conversion rates obtained from "*2001 Energy Efficiency Standards for Residential and Non-residential Buildings, California Energy Commission*," June 2001: 1 kWh = 10,239 Btu source energy; 1 Therm = 100,000 Btu source energy. 1 MMBtu =1,000,000 Btu.

[27] For measures that retain their first year savings over their entire measure life, lifecycle estimates of impacts are calculated as the first year savings times the years of effective useful life (EUL).

For dual baseline and early retirement measures, lifecycle estimates of impacts are calculated as the savings relative to the first baseline times the years of remaining useful life (RUL), plus the savings relative to the second baseline times the years of measure life after the RUL period has elapsed (EUL minus RUL).

Thus there are two factors (and any combinations of these) that may cause lifecycle GRRs to differ from first year GRRs: (1) an ex-post *impact estimate* that differs from the ex-ante impact, including any dual baseline differences, and (2) an ex-post *measure life* that is different from the ex-ante measure life, including any RUL differences.

Other useful references and appendices to this report include the following:

- Appendix C, M&V Procedures
- Appendix D, Guidance Provided with M&V Assignments on EAR Overlap and ISP
- Evaluation Guidance for Site Specific Analysis, update dated September 18, 2014[28]
- Approved List of ISP Studies[29]

3.1 Project-Specific Gross Impact Summary

Gross impact evaluation results are supported by 189 M&V sample points. A sample point is defined as one or more tracking system record representing measures that were installed at the same site under the same ProjectID or ApplicationCode. These 189 sample points are referred to in this section as "projects." Some gross impact points include only ex-ante electric savings, some include only ex-ante gas savings, and some include both ex-ante electric and gas savings. Since MMBtu is the energy metric used for the 2013 evaluation, the report does not differentiate between electric and gas results, but rather presents all results as MMBtu results. Demand savings (KW), where claimed, were analyzed and are reported separately. The original sample design called for 190 main gross M&V points targeted for completion. However, given customer willingness to participate and other factors, the final gross sample consisted of 181 main points and 8 backup points, collectively representing 41% of ex-ante MMBtu claims across all PAs (Table 3-1). Backup points serve to fill-in for main points where completion of main points is not possible.

Table 3-1: Custom Evaluation Gross M&V Sample Size Summary by PA

PA	Completed M&V Points (n)		Percent of Ex-Ante MMBtu Claims
	Main	Backup	M&V Sample
PG&E	53	2	33%
SCE	52	1	31%
SDG&E	41	2	48%
SCG	35	3	76%
All PAs	**189**		**41%**

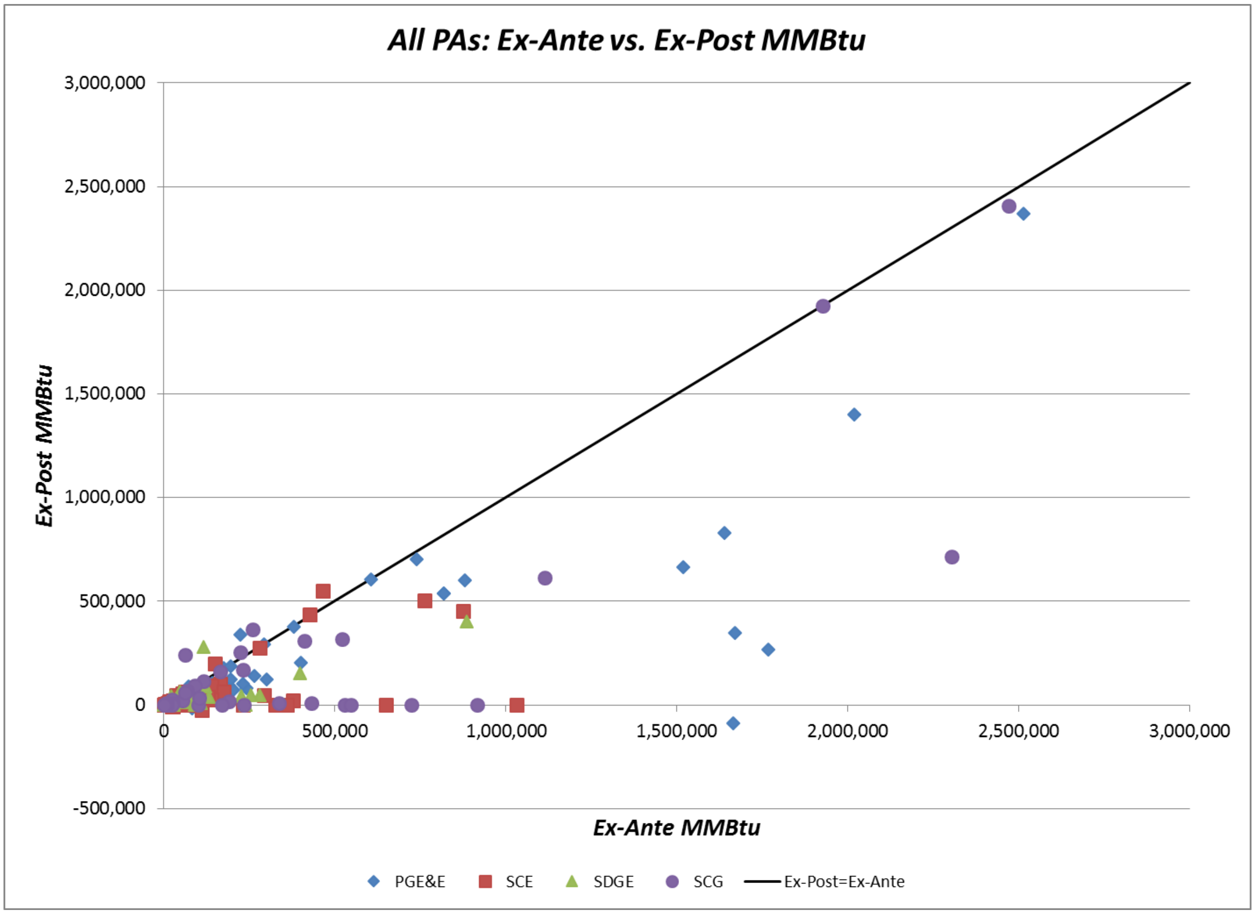
[28] Industrial, Ag and Large Commercial Evaluation Guidance available at www.energydataweb.com/cpuc/. Select the search tab, and from the drop down menus, select Portfolio Cycle 2013-2014 and Work Order (ED_I_IAL_2-Itron) 1314 IALC Impact.

[29] http://www.cpuc.ca.gov/PUC/energy/Energy+Efficiency/Ex+Ante+Review+Custom+Process+Guidance+Documents.htm

Figure 3-1 graphically displays MMBtu ex-post versus ex-ante *lifecycle* savings estimates for the statewide sample. The figure compares the ex-ante (tracking system) MMBtu savings[30] with the ex-post evaluated MMBtu savings for the M&V sample points. The chart also includes a unity line, which divides the results into those in which the project-specific realization rates were above 1.0 (sites above the line) and below one (sites below the line). All 189 projects are included in the figure.[31] PA-specific plots are included in Appendix A.

Very few of the sampled projects yielded lifecycle GRRs that exceed 1.0 and lie above the unity line in the chart.

Figure 3-1: Lifecycle Ex-Ante and Ex-Post Combined Electric and Gas Savings (MMBtu) for Sampled Projects



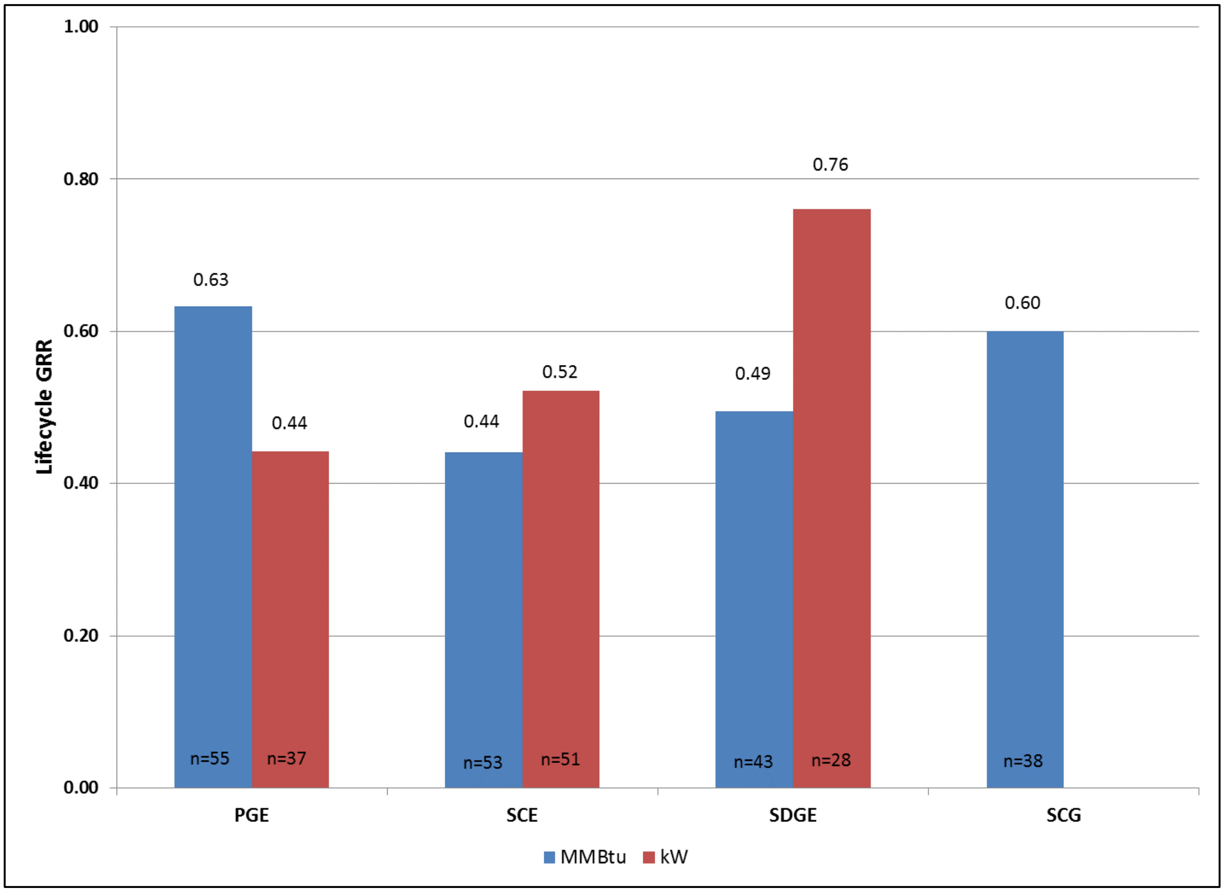
30 This figure compares "engineering estimates" for both ex-ante MMBtu and ex-post MMBtu. That is, if the PA-claimed ex-ante savings for a record include the PA RR=0.9 adjustment, that adjustment was removed for the purpose of this comparison.

31 No extreme lifecycle GRRs were found in the 2013 evaluation; this report therefore does not include discussions of results generated with- and without extreme points.

3.2 PA Gross Realization Rate Results

Weighted gross realization rates for all PAs for the appropriate energy metrics (MMBtu or kW) are presented graphically in Figure 3-2.

Figure 3-2: Lifecycle Gross Realization Rate Results by PA for Combined Electric and Gas Savings (MMBtu) and for Peak Electric Demand (kW)



As shown in the tables that follow, weighted average GRRs by PA fuel domain tend to be statistically significantly less than one and greater than zero.

Table 3-2 presents project lifecycle (LC) GRRs for each of the four PAs. The mean weighted realization rate is shown for MMBtu and kW as a separate row for each PA 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 error ratio.[32] In addition, first year (FY) GRRs are presented for comparison purposes.

Table 3-2: Weighted Project Lifecycle and First Year Gross Realization Rates by PA and Energy Metric (MMBtu and kW)

Energy Metric	Sample Count	Population Count	LC Mean GRR	Error Ratio	90% Confidence Interval	LC GRR >1.5	LC GRR =0	LC GRR <0	FY Mean GRR
PG&E									
MMBtu*	55	1,125	0.63	0.47	0.57 to 0.70	1	3	2	0.74
kW	37	853	0.44	1.41	0.28 to 0.61	1	1	1	0.54
SCE									
MMBtu*	53	934	0.44	1.07	0.34 to 0.54	1	9	3	0.54
kW	51	838	0.52	0.80	0.43 to 0.62	4	14	3	0.64
SDGE									
MMBtu*	43	264	0.49	0.88	0.40 to 0.59	4	2	1	0.75
kW	28	144	0.76	0.90	0.57 to 0.95	3	2	0	1.02
SCG									
MMBtu	38	158	0.60	0.86	0.48 to 0.72	1	11	0	0.69

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

The error ratio for most domains is similar to the error ratio obtained in the *2010-2012 WO033 Custom Impact Evaluation Final Report.*[33] In the 2013 sample, the relatively high standard deviation (e.g. MMBtu GRR for SCE) is indicative of the variability in the data rather than of a small sample size. The underlying sample has individual gross realization rates that are widely dispersed between zero and values exceeding 1.0, which will always result in a large standard deviation regardless of the number of projects sampled.

The mean lifecycle realization rates by PA and energy metric are less than 0.70 for all but one energy metric (the GRR for SDG&E kW is 0.76) and are similar to those from 2010-2012. As a comparison to results presented in the table above, Table 3-3, Figure 3-3 and Figure 3-4 present LC GRR evaluation results from the 2010-12 cycle on a combined MMBtu basis.[34] While none of the LCC GRR results (kW and MMBtu) are statistically different at the 90% confidence level

[32] The error ratio is a measure of the statistical variation in the Gross Realization Rates.

[33] http://www.calmac.org/publications/2010-12_WO033_Custom_Impact_Eval_Report_Final.pdf

[34] In the 2010-12 cycle, sampling and analysis was originally performed by fuel for each PA. While the sample design was not on a combined MMBtu basis, the 2010-12 results have been weighted by MMBtu in order to support a direct comparison with 2013 results.

between the two evaluation cycles, 2013 GRR results are lower for both energy metrics for all PAs except PG&E (the MMBtu GRR for both evaluation cycles is 0.63).

Table 3-3: 2010-2012 Weighted Project Lifecycle Realization Rates by PA and Energy Metric (MMBtu and kW)

Energy Metric	2010-2012 Mean LC Gross Realization Rate	2010-2012 90% Confidence Interval
PG&E		
MMBtu*	0.63	0.57 to 0.69
kW	0.46	0.35 to 0.58
SCE		
MMBtu	0.61	0.51 to 0.71
kW	0.57	0.47 to 0.67
SDGE		
MMBtu*	0.57	0.47 to 0.67
kW	0.82	0.46 to 1.17
SCG		
MMBtu	0.71	0.58 to 0.84

* The sample for 2010-12 was *not* designed and selected based on MMBtu.

Figure 3-3: Comparison of 2010-12 and 2013 Weighted MMBtu LC GRR Results

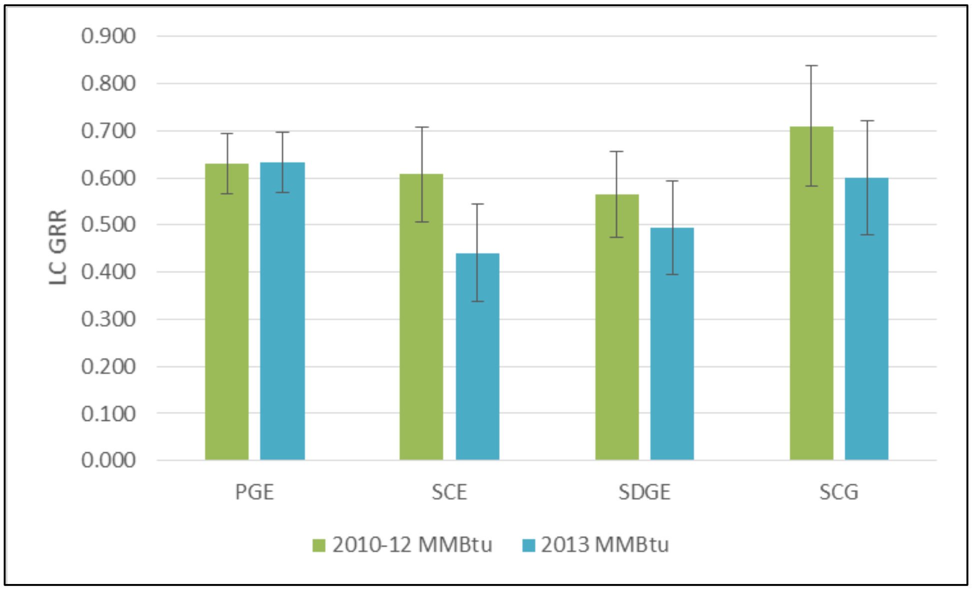
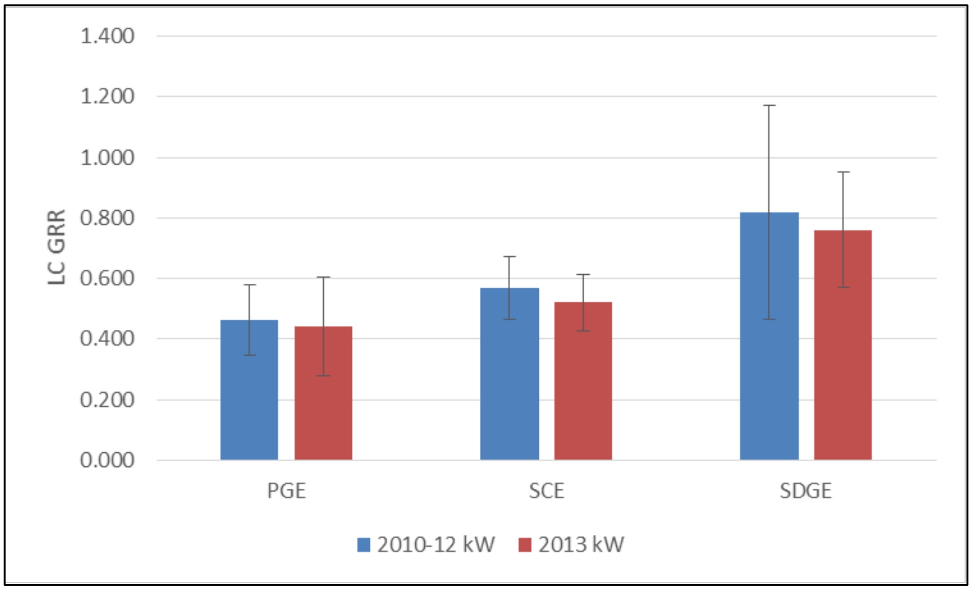


Figure 3-4: Comparison of 2010-12 and 2013 Weighted kW LC GRR Results



All 2013 weighted lifecycle realization rates are lower than the corresponding first year realization rate, which is primarily due to adjustments in measure effective useful life. In particular, of the 189 projects sampled for the 2013 evaluation, the ex-ante effective useful life (EUL) was overstated in the tracking extract for 71 measures (for example: a measure with 5-year life expectancy was assigned a 15-year EUL.) There were also 16 sampled projects claiming understated EULs, but the upward adjustments for these cases were less significant (for example: a measure with a 12-year life expectancy claiming a 10-year EUL.) Section 5.3.6 in Chapter 5 contains a more thorough description of EUL differences at the measure level.

All PAs had projects with negative and/or zero GRRs, and these served to lower the weighted realization rates. The discrepancy factors that brought about the lower realization rates are explored in the next section.

As discussed in the *IALC Research Plan*, project size was used to draw sampling strata boundaries for each PA. 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. Each PA domain has five strata, defined based on the size of claimed ex-ante MMBtu, 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. Table 3-4 presents impact results by size strata for each PA. Please note that the sample sizes for each stratum are small, and thus the stratum level results should be interpreted with caution; however, the results are illustrative of project size-related trends within the various sample domains.

Table 3-4: Project Lifecycle Realization Rates by Strata and Sample Domain (All Completed Sample Points)

PA Domain	Strata	Project Count	Projects with kW Ex-Ante	Weighted LC GRR kW	Weighted LC GRR MMBtu	LC GRR > 150%	LC GRR = 0%	LC GRR < 0%
PG&E	1	7	2	0.12	0.45	0	0	1
	2	7	5	0.98	0.76	0	0	0
	3	13	10	0.56	0.70	1	0	0
	4	13	8	0.47	0.46	0	2	1
	5	15	12	0.34	0.78	0	1	0
SCE	1	6	6	0.50	0.46	0	2	0
	2	9	9	0.17	0.24	0	4	0
	3	12	10	0.49	0.43	0	1	1
	4	12	12	0.51	0.45	0	1	1
	5	14	14	0.77	0.62	1	1	1
SDG&E	1	2	2	0.50	0.43	0	0	0
	2	3	2	0.23	0.18	0	0	0
	3	9	5	0.43	0.52	1	1	0
	4	13	5	0.89	0.44	0	1	1
	5	16	14	1.38	0.88	3	0	0
SCG	1	2	-	-	0.99	0	0	0
	2	2	-	-	0.39	0	0	0
	3	7	-	-	0.23	0	4	0
	4	11	-	-	0.39	0	3	0
	5	16	-	-	1.01	1	4	0

Observations on Table 3-4 include the following:

- Stratum-level weighted MMBtu GRRs are lower than the PA MMBtu average when that stratum has a high number of projects with negative or zero weighted GRRs (e.g. SCE's stratum 2). Similarly, a stratum containing large number of projects with GRRs higher than 1.0 will have a stratum-level GRR that is higher than the PA average (e.g. SDGE's stratum 5).
- The realization rate for kW in the PG&E electric domain, stratum 1, is 0.12; this is significantly different than the MMBtu realization rate of 0.45 for the same stratum. The cause for this divergence is the fact that stratum 1 consists mainly of gas-only projects, with only two electric projects contributing to the kW realization rate calculation; both of these have very low individual realization rates.

Given that just a few projects might be large enough to command an entire stratum, and that each stratum has roughly equal weight in the result, these observations illustrate the importance that a single project can have on the overall PA-level GRR result. There is clearly a need for the PAs to improve in the areas of estimation accuracy and quality control for all projects, but in particular there is a need to focus on projects where the ex-post savings are zero or even negative. As will be demonstrated below in Table 3-5 these projects with zero or negative savings are due primarily to three factors – baseline selection, lack of eligibility screening, and calculation methods, with the latter leading to the negative results. Baseline selection and eligibility screening are pretty basic steps in the development of ex-ante savings estimates and represent relatively easy-to-implement areas for improvement.

A summary of project-specific results for each individual gross impact project is provided in Appendix A. This appendix includes anonymized site and record identifiers, the strata, ex-ante savings claims from the PA tracking systems, gross realization rates, and the net to gross ratio.

3.3 EAR Overlap Sensitivity Analysis

The establishment of ex-ante review (EAR) is discussed in CPUC Decision 09-09-047,[35] which requires the Energy Division (ED) to review and approve ex-ante impact estimation approaches and ex-ante savings for non-DEER ("custom") measures. The ex-ante review process is designed to provide constructive early feedback to the PAs and third-party implementers, and ultimately to improve the accuracy of ex-ante savings estimation and to create a greater awareness and compliance with the CPUC policies and expectations for project documentation. All projects that have been subject to ex-ante review, and that are subsequently installed, can be fully claimed by the PAs (in other words: PA RR=1.0). To claim all other non-deemed projects, PAs adjust ex-ante estimates by a PA RR=0.9.

A total of 56 EAR projects (PA RR=1.0) were part of the IALC 2013 population: 38 were installed in PG&E territory, 12 in SCE territory, four in SDG&E territory, and two in SCG territory. The IALC stratified random sampling process selected seven EAR projects for evaluation: three from PG&E, one from SCE, two from SDG&E, and one from SCG. Table 3-5 shows the first year and the lifecycle MMBtu realization rate results for these seven EAR-reviewed points.

In order to assess the effect of the EAR process on PA-level weighted GRRs, a sensitivity analysis was performed by removing the EAR projects from both the IALC sample and the population of projects. The resulting weighted GRRs for all *custom, not EAR-reviewed, sample points* for each PA are also shown in the table for comparison purpose.[36]

35 The decision may be found at the following web link: http://docs.cpuc.ca.gov/published/FINAL_DECISION/139858.htm

36 Note that this constitutes a comprehensive sensitivity analysis only if all EAR-reviewed projects can be identified in the database by searching for IOU RR=1.0. For the purpose of this analysis, any projects identified as IOU RR=0.9 were interpreted as being not EAR-reviewed points.

Table 3-5: MMBtu Realization Rates for Sampled EAR Projects

PA Domain	Project ID	PA RR	FY GRR-MMBtu	LC GRR-MMBtu
PGE	E30019	1.00	1.00	1.00
	E30025	1.00	1.00	0.51
	E30069	1.00	1.50	1.52
	Remaining Points (n=52)	0.90	0.72	0.62
	All Sample (n=55)	0.90	0.74	0.63
	Percent change due to EAR points	1%	2%	3%
SCE	F30004	1.00	1.02	1.02
	Remaining Points (n=52)	0.90	0.53	0.43
	All Sample (n=53)	0.90	0.54	0.44
	Percent change due to EAR points	0%	3%	3%
SDGE	H30016	1.00	0.94	0.69
	H30019	1.00	1.00	0.75
	Remaining Points (n=41)	0.90	0.74	0.48
	All Sample (n=43)	0.90	0.75	0.49
	Percent change due to EAR points	0%	1%	2%
SCG	G30001	1.00	1.00	1.00
	Remaining Points (n=37)	0.90	0.64	0.56
	All Sample (n=38)	0.91	0.69	0.60
	Percent change due to EAR points	1%	7%	8%

Observations for Table 3-5:

- Six of the seven EAR projects sampled have first year GRRs of 1.0 or higher, and the remaining project has a first year GRR of 0.94. For all PAs the weighted first year GRRs of the remaining (non-EAR) projects are lower than the GRRs for EAR projects.
- Four of the seven EAR projects have lifecycle GRRs of 1.0 or higher. For the other three EAR projects the evaluation found significantly shorter EULs than PA ex-ante EULs specified in the tracking extract. This caused these three lifecycle GRRs to be lower than their first year GRRs, and in some cases (E30025) also lower than the weighted lifecycle GRR of the remaining (non-EAR) projects.

Table 3-5 also shows a comparison between the PA-level weighted MMBtu GRR for all sampled projects, and the percent change in MMBtu GRR that can be attributed to including the EAR projects into the sample. The presence of EAR projects in the 2013 IALC population (and sample) improves the first year weighted GRR by one to seven percent, and the lifecycle weighted GRR by two to eight percent for all PAs.

3.4 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 analysis is based on discrepancies associated with first year MMBtu impacts.[37]

When gross ex-post impacts for a sampled project were found to be different than the PA ex-ante impacts, the evaluation documented the associated discrepancy factors. For some projects there was only one factor (e.g. the PA calculation method was not appropriate, and another, more appropriate method was used for the evaluation) while for others there were multiple factors (e.g. ex-post operating hours observed in the field were different than the number of hours documented in project paperwork *and* the number of measures installed was also different than that reported). Ultimately, individual discrepancy factors were classified into seven important categories: operating conditions, calculation method, inappropriate baseline, ineligible measure, inoperable measure, measure count, and tracking database discrepancy.[38] When examined for both the frequency of occurrence and the degree of impact on the ex-ante savings claims, the following four factors are most influential (or dominant):[39]

- Differences in operating conditions (for example, changes in hours of operation, VSD speeds, return to original operation, changes in production levels, etc.).[40]
- Calculation methods used for ex-post savings estimation were different than those used to estimate ex-ante savings. Some examples include: using AIRMaster+_ model instead of a spreadsheet model; switching from Energy Pro to eQUEST building simulation software; different engineering calculation approaches based on post-retrofit or post-construction data availability; disallowing load forecasting; use of expanded spreadsheet approaches to account for varying loads and interactive effects; 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; and etc.
- Inappropriate baselines or baseline conditions used for ex-ante savings estimation. Some examples of baseline-related issues are: rejected early replacement claims; new equipment

[37] The effect of ex-post dual baseline adjustments on lifecycle GRRs is not reflected in this discrepancy analysis.

[38] A separate 'Other' category includes less common factors and accounts for a relatively small number of projects and percentage of savings. These factors are program rule compliance issues, measure not installed, and unquantified fuel impacts.

[39] Please note that, while inappropriate baseline may ordinarily cause a downward adjustment (ex-post lower than ex-ante impacts), adjustments to the operating conditions and/or calculation method sometimes cause an upward adjustment (ex-post higher than ex-ante impacts).

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

that does not exceed code-required efficiency levels; new equipment that does not exceed industry standard efficiency levels; rejected normal replacement claims; inaccurate baseline or pre-retrofit operating hours; and etc.

- Ineligible measures were another primary reason for downward adjustment of the ex-ante MMBtu impacts. Some examples surrounding ineligible measures include the following: program rules that do not allow repairs; like-for-like replacements; retrofit measures that did not exceed codes and industry standard practices (ISP); other program rule violations, and operational changes (such as HVAC control measures involving temperature changes).

3.4.1 Summary of Discrepancy Factor Impact

Given multiple tracking records associated with some projects, 240 records associated with the impact sample of 189 projects were examined (3.5 Million MMBtu ex-ante impacts). For 42 records, the evaluation found no discrepancies (0.5 MMBtu ex-ante impacts were not adjusted.) For the balance of 198 records, ex-post impacts were different from ex-ante MMBtu impacts; 157 records, affecting 2.5 MMBtu ex-ante impacts, were adjusted downward, and 41 records, affecting 0.5 MMBtu ex-ante impacts, were adjusted upward. A summary of these adjustments is presented in this section.

3.4.2 Discrepancy Factor Assessment for Projects with the Lowest GRRs

A very important subgroup of records is that corresponding to sampled projects with zero or negative MMBtu GRRs. There were 25 projects for which the ex-post MMBtu impacts were zero, and six for which the ex-post impacts for the project were negative. Table 3-6 identifies these projects and the factors that led to the zero or negative ex-post MMBtu impacts.

For projects with zero ex-post MMBtu, the discrepancy factors that occur most frequently are inappropriate baseline and ineligible measure. Note that either of these would cause a zero GRR and that some projects that appear here under the inappropriate baseline heading ultimately lead to measure ineligibility and were analyzed as such in Chapter 5 (PPA analysis). Calculation method was the factor that occurred most frequently for projects with negative ex-post MMBtu.

Table 3-6: Discrepancy Factors for Projects with Zero or Negative Ex-Post MMBtu

ItronID	First Year MMBtu GRR	Lifecycle MMBtu GRR	Customer Agreement Date	Discrepancy Factor				
				Inappropriate Baseline	Ineligible Measure	Calculation Method	Inoperable Measure	Operating Conditions
E30140	0	0	3/11/2013	X				
E30176	0	0	10/2/2012	X	X			
E31035	0	0	9/5/2012					X
F30003	0	0	9/7/2012	X				
F30006	0	0	5/16/2013	X				
F30014	0	0	3/11/2011		X			
F30015	0	0	8/10/2012	X				
F30024	0	0	8/15/2012		X			
F30026	0	0	9/26/2012	X				
F30042	0	0	10/30/2012	X				
F30093	0	0	9/7/2012		X			
F30922	0	0	9/18/2012				X	
G30005	0	0	11/7/2013	X			X	
G30006	0	0	9/22/2011	X				
G30007	0	0	4/5/2012				X	
G30010	0	0	2/28/2012	X				
G30016	0	0	6/24/2013		X			
G30021	0	0	3/5/2013	X				
G30028	0	0	12/17/2012	X				
G30066	0	0	3/7/2013	X				
G30072	0	0	7/28/2010			X		
G30088	0	0	11/7/2012	X				
G30095	0	0	11/7/2012	X				
H30009	0	0	11/8/2012		X			
H30029	0	0	12/3/2012		X			
E30006	-0.08	-0.05	5/24/2011		X	X		
E30170	-0.19	-0.19	5/23/2013			X		
F30049	-0.34	-0.22	11/2/2012			X		X
F30193	-0.28	-0.28	9/24/2013			X		
F30249	-0.36	-0.24	11/13/2012			X		
H30046	-0.06	-0.05	12/3/2012			X		
Discrepancy Factor Frequency				15	8	7	3	2

* None of these projects listed Measure Count, Tracking Discrepancy or Other as ex-post discrepancy factors.

3.4.3 Assessment of Downward and Upward Adjustments to Ex-Ante Claims by Discrepancy Factor

As described above, each record was assigned a primary (and sometimes a secondary and tertiary) factor that explains the observed discrepancy in ex-post vs. ex-ante estimates. The fraction of the discrepancy attributable to each factor was also recorded. Table 3-7 summarizes the downward adjustments by discrepancy factor (including projects with zero and negative ex-post MMBtu,) and Table 3-8 summarizes the upward adjustments to ex-ante, by discrepancy factor.[41] Figure 3-5 displays the same information contained in Table 3-7 and Table 3-8 for all PAs combined. PA-specific plots can be found in Appendix A.

Individual discrepancy factors caused different downward adjustments; for example, at the statewide level: inappropriate baseline caused a 79 percent reduction, calculation method caused a 51 percent reduction, and operating conditions caused a 28 percent reduction in the ex-ante impacts of affected records. Collectively, downward adjustments affected 157 records and caused a 49 percent reduction of the 2.5 Million MMBtu ex-ante impacts; this is equivalent to an unweighted realization rate of 0.51 for these 157 records.

All factors combined for this set of applicable records, SCE and SDG&E claims were most affected by these downward adjustments, at 66 percent and 57 percent, respectively, versus 49 percent across all PAs. PG&E and SCG have smaller, but still significant, downward claims adjustments of 43 and 44 percent.

As stated above, upward adjustments affected a total of 41 records and caused a 29 percent increase to the 0.5 Million MMBtu ex-ante impact claims; this is equivalent to an unweighted realization rate of 1.29 for these 41 records. Only two of the factors, operating conditions and calculation method, were found to commonly lead to upward adjustments to ex-ante savings estimates. Sample sizes for other factors are too small to warrant further discussion.

All factors combined for this set of applicable records, PG&E and SCG claims were most affected by these upward adjustments, at 40 percent and 30 percent, respectively, versus 29 percent across all PAs. Both SDG&E and SCE have smaller upward claims adjustments of 19 percent and 11 percent, respectively.

[41] If a given record was adjusted as a result of more than one discrepancy factor, that record is counted under each of the discrepancy factors shown in the table. MMBtu was not double-counted; rather, the fraction of the adjustment attributable to each factor was used to break down the record's ex-ante and ex-post MMBtu into the reason-related components shown.

Figure 3-5: Ex-post Upward and Downward Adjustments to Ex-ante MMBtu for Sampled Projects - All PAs

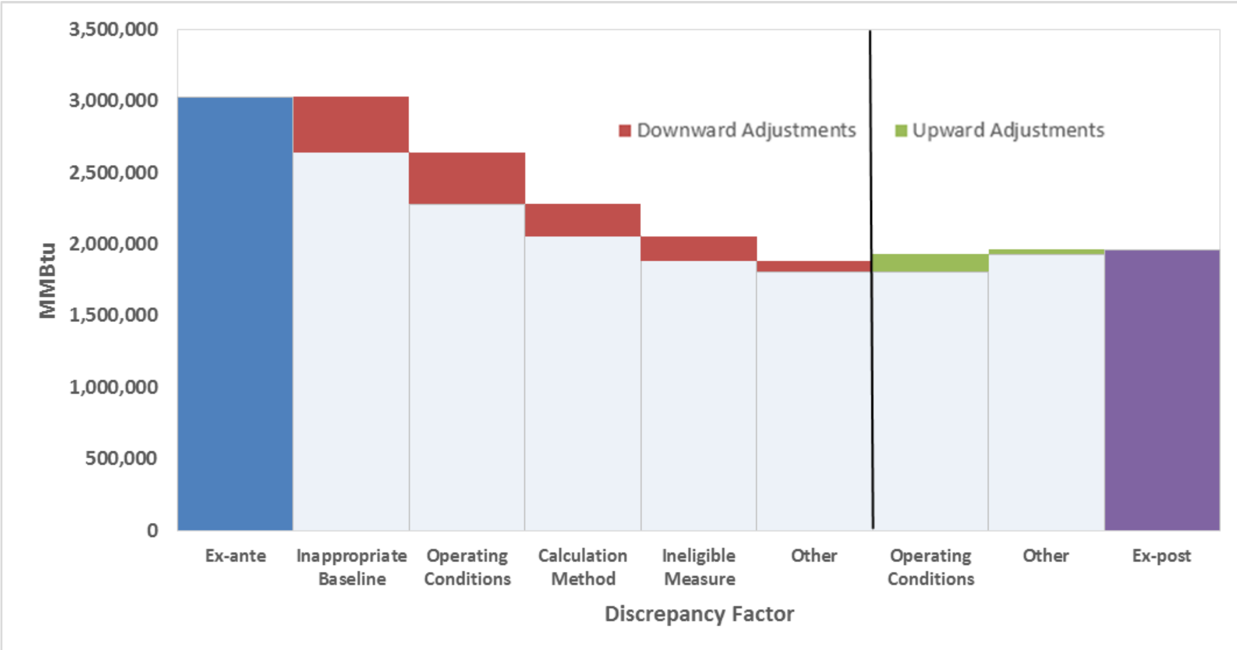


Table 3-7: Records with Ex-Post Downward Adjustments to Ex-Ante MMBtu Impacts, by Discrepancy Factor; Statewide and by PA

Discrepancy Factor	n Records	Ex-Ante MMBtu Affected	Ex-Post MMBtu Change	% of Ex-Post MMBtu Change	% Ex-Ante Reduction
All PAs					
Inappropriate Baseline	30	493,575	-389,436	32%	-79%
Operating Conditions	48	1,291,890	-360,511	29%	-28%
Calculation Method	72	443,068	-225,627	18%	-51%
Ineligible Measure	14	182,899	-171,455	14%	-94%
Inoperable Measure	4	88,898	-75,161	6%	-85%
Measure Count	2	4,786	-822	0%	-17%
Tracking Discrepancy	5	153	-7	0%	-5%
Other	1	653	-376	0%	-58%
All Records	157	2,505,923	-1,223,396	100%	-49%
PGE					
Operating Conditions	18	628,720	-211,943	47%	-34%
Ineligible Measure	4	85,504	-85,504	19%	-100%
Calculation Method	17	143,592	-82,062	18%	-57%
Inappropriate Baseline	9	164,924	-63,450	14%	-38%
Inoperable Measure	1	16,745	-3,008	1%	-18%
Measure Count	1	4,735	-788	0%	-17%
Tracking Discrepancy	0	0	0	0%	0%
Other	0	0	0	0%	0%
All PGE Records	43	1,044,221	-446,754	100%	-43%
SCE					
Inappropriate Baseline	8	183,395	-181,433	54%	-99%
Calculation Method	17	147,591	-68,449	20%	-46%
Ineligible Measure	3	44,338	-44,338	13%	-100%
Operating Conditions	14	138,980	-44,289	13%	-32%
Inoperable Measure	1	15	-15	0%	-100%
Other	1	653	-376	0%	-58%
Measure Count	0	0	0	0%	0%
Tracking Discrepancy	0	0	0	0%	0%
All SCE Records	41	514,973	-338,899	100%	-66%
SDGE					
Calculation Method	30	102,882	-56,996	68%	-55%
Ineligible Measure	5	27,789	-19,898	24%	-72%
Operating Conditions	7	13,557	-4,436	5%	-33%
Inappropriate Baseline	2	2,908	-2,358	3%	-81%
Tracking Discrepancy	5	153	-7	0%	-5%
Measure Count	1	51	-35	0%	-68%
Inoperable Measure	0	0	0	0%	0%
Other	0	0	0	0%	0%
All SDGE Records	30	147,340	-83,731	100%	-57%
SCG					
Inappropriate Baseline	11	142,348	-142,195	40%	-100%
Operating Conditions	9	510,633	-99,844	28%	-20%
Inoperable Measure	2	72,138	-72,138	20%	-100%
Calculation Method	8	49,003	-18,120	5%	-37%
Ineligible Measure	2	25,268	-21,715	6%	-86%
Measure Count	0	0	0	0%	0%
Tracking Discrepancy	0	0	0	0%	0%
Other	0	0	0	0%	0%
All SCG Records	43	799,389	-354,012	100%	-44%

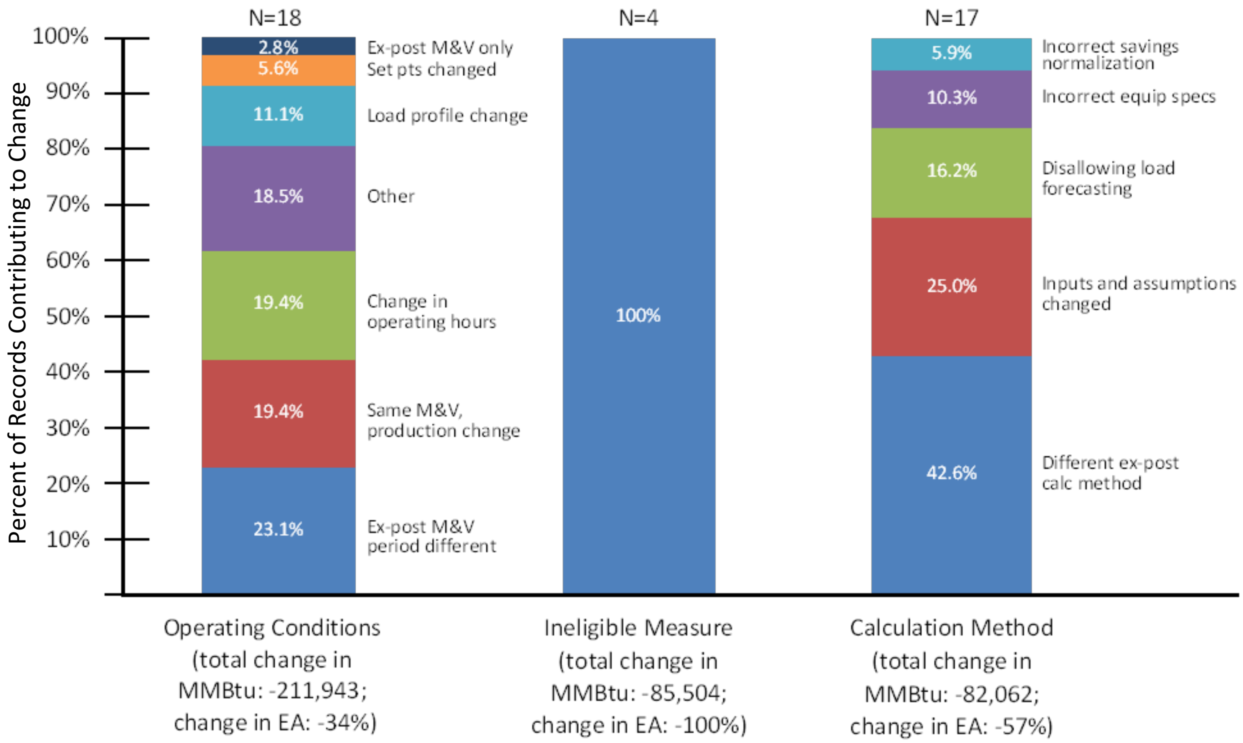
Table 3-8: Records with Ex-Post Upward Adjustments to Ex-Ante MMBtu Impacts, by Discrepancy Factor; Statewide and by PA

Discrepancy Factor	n Records	Ex-Ante MMBtu Affected	Ex-Post MMBtu Change	% of Ex-Post MMBtu Change	% Ex-Ante Increase
All PAs					
Operating Conditions	21	439,042	120,119	78%	27%
Calculation Method	20	65,723	18,817	12%	29%
Tracking Discrepancy	1	5,767	12,776	8%	222%
Measure Count	1	1,464	624	0%	43%
Other	1	12,195	1,659	1%	14%
Inappropriate Baseline	0	0	0	0%	0%
Ineligible Measure	0	0	0	0%	0%
Inoperable Measure	0	0	0	0%	0%
All Records	41	524,191	153,995	100%	29%
PGE					
Operating Conditions	6	273,082	110,673	100%	41%
Calculation Method	5	5,411	441	0%	8%
Inappropriate Baseline	0	0	0	0%	0%
Ineligible Measure	0	0	0	0%	0%
Inoperable Measure	0	0	0	0%	0%
Measure Count	0	0	0	0%	0%
Tracking Discrepancy	0	0	0	0%	0%
Other	0	0	0	0%	0%
All PGE Records	11	278,493	111,114	100%	40%
SCE					
Calculation Method	4	18,820	6,486	65%	34%
Operating Conditions	5	75,125	3,467	35%	5%
Inappropriate Baseline	0	0	0	0%	0%
Ineligible Measure	0	0	0	0%	0%
Inoperable Measure	0	0	0	0%	0%
Measure Count	0	0	0	0%	0%
Tracking Discrepancy	0	0	0	0%	0%
Other	0	0	0	0%	0%
All SCE Records	9	93,945	9,953	100%	11%
SDGE					
Tracking Discrepancy	1	5,767	12,776	60%	222%
Operating Conditions	8	88,134	5,613	26%	6%
Calculation Method	8	17,185	2,270	11%	13%
Measure Count	1	1,464	624	3%	43%
Inappropriate Baseline	0	0	0	0%	0%
Ineligible Measure	0	0	0	0%	0%
Inoperable Measure	0	0	0	0%	0%
Other	0	0	0	0%	0%
All SDGE Records	5	112,550	21,283	100%	19%
SCG					
Calculation Method	3	24,307	9,620	83%	40%
Operating Conditions	2	2,701	366	3%	14%
Other	1	12,195	1,659	14%	14%
Inappropriate Baseline	0	0	0	0%	0%
Ineligible Measure	0	0	0	0%	0%
Inoperable Measure	0	0	0	0%	0%
Measure Count	0	0	0	0%	0%
Tracking Discrepancy	0	0	0	0%	0%
All SCG Records	16	39,203	11,645	100%	30%

3.4.4 Categorical Explanation for Primary Discrepancy Factors

The discrepancy factors that correspond to the largest downward adjustments for each PA are examined in detail in this section. During ex-post evaluation activities, further explanatory categories were listed with each discrepancy factor, and these sub-categories are presented in this section of the report. Figure 3-6 addresses the factors that cause the three largest downward adjustments to ex-ante MMBtu for PG&E and provides the frequency with which each sub-category was noted for each of these primary factors.

Figure 3-6: Most Influential Discrepancy Factors that Caused Downward Adjustments for PG&E

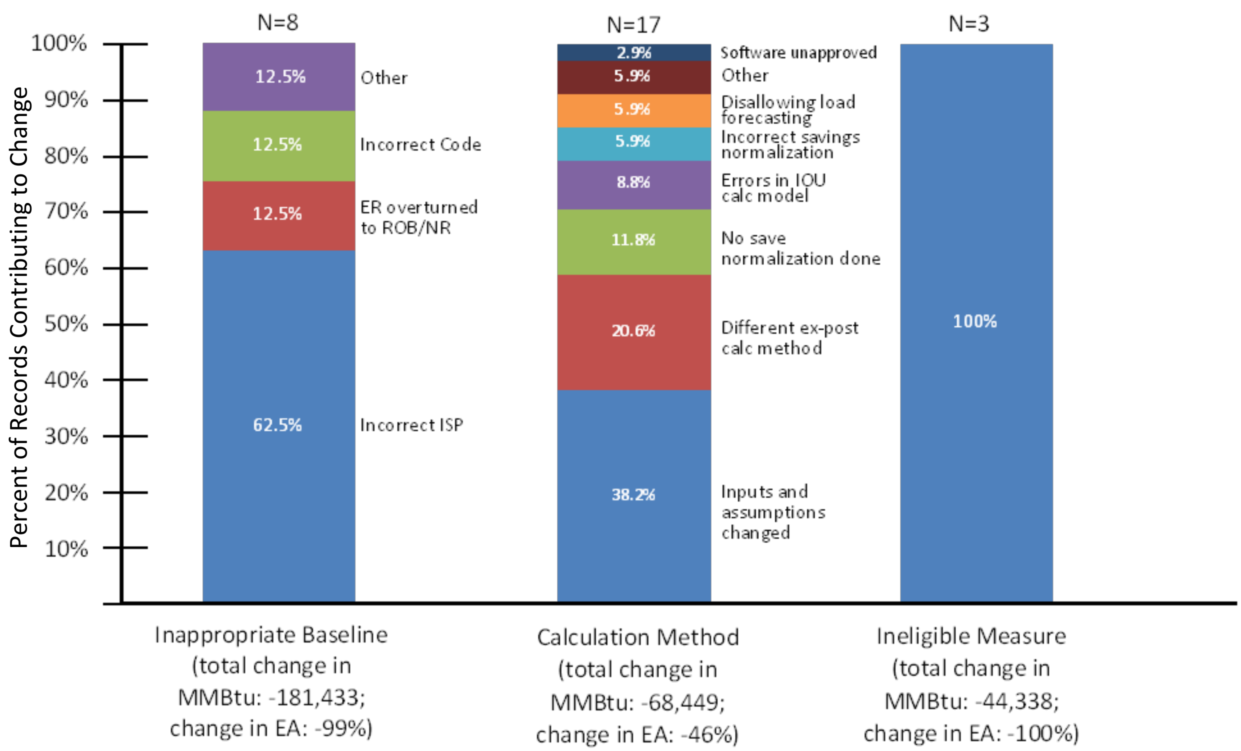


For PG&E the top three discrepancy factors that resulted in a downward adjustment of ex-ante MMBtu impacts were operating conditions, ineligible measure, and calculation methods. Changes in operating conditions affected 18 records and resulted in the highest relative reduction of the ex-ante MMBtu claims, representing a 34 percent overall reduction of the claimed 628,720 MMBtu for this category. Use of longer ex-post M&V periods by the evaluation was the most frequently observed sub-category within the operating conditions discrepancy group, while changes in verified production levels and operating hours represented the second and third most frequent contributing factors to the reduction of ex-ante MMBtu claims.

The evaluation found ineligible measures to be the second reason for downward adjustment to PG&E's ex-ante MMBtu claims. Although, measure ineligibility affected only four records, it resulted in a significant reduction in the overall ex-ante claims. Measures were deemed ineligible for incentives when the PA-supplied fuel was not impacted by the measure, or if the measure entailed a like-for-like replacement.

Calculation method changes during the ex-post analysis resulted in the third highest ex-ante MMBtu claim reduction and represented a 57 percent reduction of the 143,592 MMBtu claim. Seventeen records were affected. Of these, seven records had different ex-post engineering calculation approaches; for four records the evaluation changed PA inputs and assumptions; and for three records the evaluation disallowed load forecasting.

Figure 3-7: Most Influential Discrepancy Factors that Caused Downward Adjustments for SCE

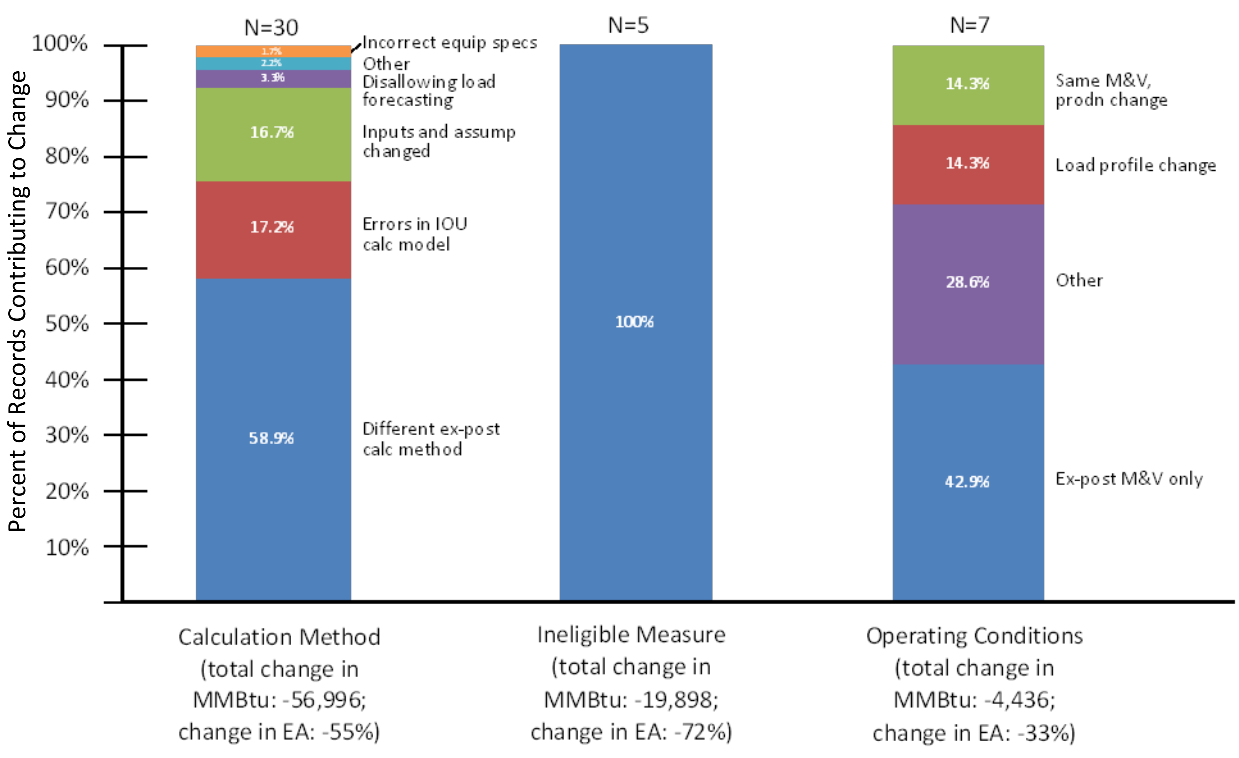


For SCE the top three discrepancy factors that resulted in a downward adjustment of ex-ante MMBtu impacts were inappropriate baseline, calculation methods and ineligible measures (Figure 3-7). Of the eight records affected by inappropriate baseline issues, five had incorrect ISP application. For one record (F30003) the evaluation overturned the PA's claim for early retirement to a replace on burnout (ROB) baseline based on site level findings. For another record (F30170) the evaluation adjusted the project baseline type to a more appropriate code baseline. In the last of the affected records the PA baseline was revised after it was determined that the ex-ante baseline had been established by using information from a different site. Inappropriate baselines and their adjustments represented a 99 percent reduction to the claimed 183,395 ex-ante MMBtu for this sub-category.

Seventeen records were affected by changes in ex-post calculation methods. For seven records the evaluation changed PA inputs and assumption. Three records had different ex-post engineering calculation approaches. For two others the PA did not normalize the savings to production levels. Changes to ex-ante calculation methods resulted in a 46 percent reduction to the claimed 147,591 MMBtu.

Ineligible measures were the third largest reason for downward adjustment to SCE's ex-ante MMBtu claims. Measures entailing like-for-like replacements were deemed ineligible for incentives, and this resulted in a 100 percent reduction of the 44,338 MMBtu ex-ante claim for these records.

Figure 3-8: Most Influential Discrepancy Factors that Caused Downward Adjustments for SDG&E

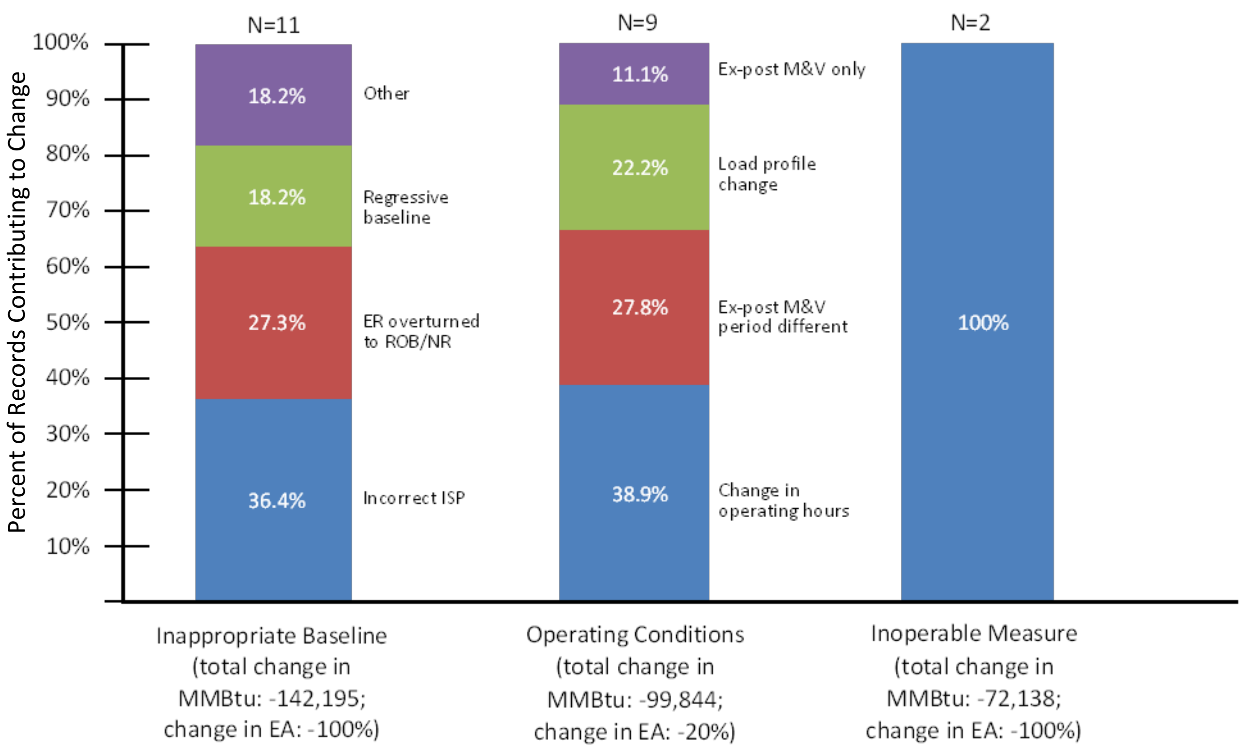


For SDG&E the top three factors that resulted in a downward adjustment of ex-ante MMBtu impacts were calculation methods, ineligible measures, and operating conditions (Figure 3-8). For 30 records calculation method changes during ex-post analysis resulted in the highest ex-ante MMBtu claim reduction and represented a 55 percent reduction of the 102,882 MMBtu claim. Changes to the ex-post engineering calculation methods was the most dominant sub-category, affecting 18 records. For five records the evaluation identified errors in the PA calculations. For another five records the evaluation updated PA inputs and assumptions.

Ineligible measures were the second largest reason for downward adjustment to SDG&E's ex-ante MMBtu claims, affecting five records. Measures entailing like-for-like replacements, reprogramming of existing controls, and measures with less than 2-year simple payback period were deemed ineligible for incentives. Measure ineligibility resulted in a 72 percent reduction to the 27,789 MMBtu associated with those projects.

Changes in operating conditions, affecting seven records, caused the third largest reason for reduction to SDG&E's ex-ante MMBtu claims. These changes represent a 33 percent reduction of the claimed 13,557 MMBtu for this category. For three records the evaluation found that there was no PA M&V conducted; for two others the observed setpoints and operating hours were found to be different from ex-ante reported values. Load profile and production level changes also contributed to this category.

Figure 3-9: Most Influential Discrepancy Factors that Caused Downward Adjustments for SCG



For SCG the top three reasons that resulted in a downward adjustment of ex-ante MMBtu impacts were inappropriate baseline, operating conditions and inoperable measures (Figure 3-9). Of the 11 records impacted by the inappropriate baseline issues, four had incorrect ISP application. For three records the evaluation overturned the PA's claim for early retirement to a normal replacement

(NR) baseline based on site-level findings. For two records, the evaluation disallowed a regressive baseline, as there was sufficient evidence that the existing equipment was above the minimum baseline requirement. Other code-based adjustments affected the remaining two records. Adjusting for inappropriate baselines resulted in a nearly 100 percent reduction of the claimed 142,348 ex-ante MMBtu for this primary discrepancy factor.

Changes in operating conditions affected nine records and resulted in the second highest reduction of the ex-ante MMBtu claims. This represents a 20 percent overall reduction to the claimed 510,633 MMBtu for this primary discrepancy factor. Change in measure operating hours was the most frequently observed sub-category within the operating conditions discrepancy group affecting four records. Longer M&V duration and changes in load profiles affected another four records and represented the second and third most frequent contributing factors to the reduction of ex-ante MMBtu claims.

Inoperable measures is the third most important factor reducing SCG's ex-ante MMBtu claims. Two records were affected due to this discrepancy factor, which entailed a 100 percent reduction of the claimed 72,138 MMBtu. In one instance the business had moved locations; in the other the measure was decommissioned and replaced. The evaluation set the ex-post savings to zero for both these records.

3.5 Evaluation Suggestions and Considerations to Address the Most Influential Discrepancy Factors

During the site impact evaluation activities, evaluation engineers provided suggestions and considerations for improving PA ex-ante savings estimates. These were integrated with the discrepancy analysis discussed in Section 3.4 Those suggestions and considerations are summarized below for each of the main discrepancy factors noted above -- operating conditions, calculation methods, inappropriate baseline and ineligible measures. Note that not all sub-categories by primary discrepancy factor received suggestions from the engineers.

3.5.1 Operating Conditions

Consistent with 2010-2012 custom evaluation results, changes in operating conditions represent the single greatest cause for evaluation-based reduction to ex-ante saving claims. Some variation and change is normally expected between the pre-installation and post-installation periods; however, there are additional steps the PAs and implementers can take to improve ex-ante savings estimates. The following are a few suggestions and considerations from the evaluation engineers to help address these issues.

Same M&V, Load Profile Change

- Verify that savings calculations are based on actual occupancy schedules and reflect the post-installation conditions accurately. Identify any changes to system operating pressures, temperatures, or flows, and adjust the savings models to the new operating conditions after ensuring that measure operation and production levels are stable (E30095, G30010).
- Ensure that the correct data are used to develop efficiency profiles in regression models; if data are adjusted later, then document what changed and why. Avoid high polynomial curve fits with low R^2 values. DOE Superior Energy Performance (SEP) M&V protocols require an R^2 value of not less than 0.5 for model testing, with a p-value ranging from 0.1 to 0.2[42] (H30060).
- For pump efficiency improvement projects, use actual pump efficiency tests, not estimates. For projects entailing weather sensitive measures, verify that the correct weather files are used in the analysis.

Ex-post M&V Period Different (Longer Term)

- The PAs should consider longer-term pre- and post-installation M&V activities and true-up the savings estimates to reflect current and representative measure operation. Additionally, the PAs should use trend data over a longer time duration to better characterize key parameters in order to perform a fair comparison of pre- and post-installation energy usage/demand (E30014, F30005, F30601, G30003, G30004, G30015, H30006).
- Adjust calculations for the post-installation discharge pressure, pump depths and fluid levels. Use a period longer than two weeks during post-retrofit M&V in conjunction with SCADA data, and the use of more conservative assumptions for pre-installation energy metrics. Also, use non-static efficiencies for pumps and motors reflecting load changes (F30028, F30049).

Production Change

- Before submitting the final savings, the PAs should normalize for production fluctuations between pre- and post-installation periods (F30007, F30018, F30601).
- Collect longer periods of production data to determine typical post-project production levels (E30005, F30049).

Changes in Operating Hours

- Provide pre- and post-installation data supporting claims of annual operating hours (G30039).

42 http://www.energy.gov/sites/prod/files/2014/07/f17/sep_mv_protocol.pdf (Section 3.4.5 pg. 10).

- The PAs should use available SCADA/ EMS data to estimate operating hours, instead of facility shift hours (E30006).
- The PAs should be more conservative when estimating annual operating hours and wait for stable measure operation before completion of M&V and impact calculations, particularly if there is reason to believe the measure-operating schedule may change. Any updates can be easily captured at the time of the installation report review. This suggestion goes with other commercial SBD recommendations from evaluators that the eQUEST or EnergyPro simulation models be re-run after the building is commissioned, more completely occupied, and in steady state operation (E30004).
- List operating hours for specific groups of equipment rather than using the facility operating hours in savings calculations (G30009).
- Conduct due diligence to ascertain the annual operating profile of equipment serving variable loads, especially with respect to seasonal variation in production. Even short-term M&V will remedy incorrect assumptions of measure annual operating hours (G30019).

No Post-Installation PA M&V, but Ex-post M&V Conducted

- True-up savings based upon post-installation data, such as by calibrating the simulation model to utility usage data (E30068, G30131).

3.5.2 Calculation Methods

Ex-Post Calculation Method Different from PA

- Provide a fully workable savings model. Carefully detail how temperature reset changes will affect energy use. Compare as much post retrofit energy data as possible to adjust claim (E30088).
- Use standardized and/or approved calculation tools to determine savings for common measures such as boilers (E30161).

Errors Found in PA Calculation Model

- Ensure that savings unrelated to the measure are not included in the ex-ante saving estimates (H30036).
- Carefully consider all inputs and coincidence of peak kW; use any post-installation M&V power data, and compare the loads during peak summer time period of 2-5pm from June 1st through September 30th (or on the actual California climate zone three day period, if data was collected during the actual three day peak for that region). Never assume continuous operation when calculating kW demand impacts and instead, require site-specific evidence to support coincident peak demand savings (H30006, H30050, H30060).
- Check EnergyPro output files for large changes in peak demand that would suggest possible anomalies in the simulation results caused by thermal lag, which delays the onset

of peak impacts by one hour resulting in very large differences between the simulated and calculated results (E30095).

Same Calculation Methods, Inputs and Assumptions Changed

- Confirm and use actual equipment specifications instead of default efficiency levels whenever possible (E30361, G30097).
- Provide a clear description of the baseline, including references and documentation sources for all values used in the savings analysis and explain any discrepancies between initial and final measured values and savings estimates (G30039, H30131).
- Use measure level data instead of building level data to isolate actual savings resulting from measures (H30014).
- Take spot measurements for voltage, power factor (PF), and kW and not just amperage. In absence of measured data, include a reasonable PF when calculating motor power, preferably based on motor specifications or nameplate. The PF for a normal motor should be in the range of 0.75 to 0.85 (H30030, H30045).

Other

- Use the most relevant data when developing load curves in addition to using specific climate-zone weather data (H30060).

Incorrect Methods Used for Saving Normalization

- For compressed air projects, use CFM to normalize energy savings not pressure or production hours (F30102).
- Normalize for weather when applicable and verify operation of equipment with post-case trends (H30027).
- For all retrofit measures, it is important to normalize production data or weather between pre- and post-installation periods. Always match the time periods, if at all possible, for instance the first 7 weeks of 2013 compared with the first 7 weeks of 2014.

No Savings Normalization Done

- Ensure that M&V captures representative operating conditions and adjust for production levels, if possible (F30049, G30025, G30048).

3.5.3 Inappropriate Baseline

Wrong ISP/code Corrected with Right ISP/code

- For all capacity expansion projects, a new ISP equipment baseline must be established that meets the post-installation operating and production capacities. In-situ equipment is an invalid baseline to calculate energy savings for capacity expansion unless it's above code or ISP (E30176).

- Do not allow RTO measure installations at manufacturers who service either automotive or aerospace industries. Recuperative units as baseline for other industries should be researched and ISP guidelines will need to be established based on current market trends (G30010, G30014).
- When considering new high efficiency equipment, incentive applications should include quotes for available new, less efficient equipment (baseline) of the same functionality to support availability (G30028).

ER Overturned to ROB/NR

- Require all applications to include the remaining useful life (RUL) of the pre-existing equipment and carefully review the assumptions involved in the calculation of RUL for all early retirement applications (in general). Correctly classify project baseline based on condition of replaced equipment (E30600, F30003, G30006, H30034).
- Conduct appropriate due diligence to insure that the current removed system would not be failing for an ER project (E30161).

Other

- The PAs should complete their assigned research, as directed from the CPUC EAR team, on the industrial boiler efficiency ISP baseline study. Until CPUC approval, use the minimum efficiency value of 82 percent (E30013, E30014).

Regressive Baseline

- PAs should push their customers into incremental energy savings over non-regressive baseline equipment. For example, customers should use thicker or better insulating materials for pool covers than were previously in use at a given site. Otherwise the installation constitutes a like-for-like replacement, which is not considered to be an energy efficiency action (E30140).
- For normal replacement measure retrofits, using an NC PA baseline study could be problematic in allowing a regressive baseline. Retrofitting with a measure that is of equivalent efficiency is not an energy efficiency action (G30016).

3.5.4 Ineligible Measure

- Ensure that electricity supplied is from the PA grid. For E30006 the electricity is supplied by a Muni, and therefore PPP charges are not collected and this project is not eligible for PA incentives.
- The PAs should ensure that the installed measures exceed code/ISP baseline performance levels and do not entail like-for-like replacements, or regressive baselines (E30161, G30016, G30046, H30029).

- Clearly document the energy efficiency action that is being performed. Provide nameplate details of pre-retrofit ESP and post-retrofit ESP, along with electrical & mechanical specifications. Ensure conformance with ISP guidelines (F30014, F30024).
- Ensure that the installed measures meet program payback requirements (H30054).

4

NTG Results

The methodology used to develop the individual, site-specific net-to-gross estimates is summarized in the IALC Research Plan.[43] Here we present the weighted results for each sampling domain. NTG, as reported here, is inclusive only of free ridership effects (1-FR) and does not include spillover or market effects.[44]

4.1 Number of Completed Surveys

One hundred and forty-six NTG surveys were completed in total. Across all four Program Administrators (PAs),[45] the number of completed surveys was roughly proportional to the population of completed projects for each domain. Table 4-1 below reports the number of completed telephone surveys by utility, including the number of main versus backup points used and the percent of ex-ante MMBtu claims represented. PG&E and SCE customer projects each accounted for roughly one-third of the completed surveys, respectively, while the remaining one-third was fairly evenly split between SDG&E and SCG projects.

Table 4-1: Completed Surveys by PA

PA	Completed NTG Points (n)		Percent of Ex-Ante MMBtu Claims
	Main	Backup	NTG Sample
PG&E	34	17	33%
SCE	30	16	29%
SDG&E	22	6	41%
SCG	17	4	60%
All PAs	**146**		**37%**

43 http://www.energydataweb.com/cpuc/search.aspx, "Custom Impact Evaluation – 2013-2014 – DRAFT RESEARCH PLAN

44 The IALC Custom NTG surveys also include a battery of questions to address participant spillover. However, these data are analyzed and reported on as part of the 2013-14 Nonresidential Spillover Study under the Residential Roadmap and the Market Studies PCG. The 2013-14 Nonresidential Spillover Study evaluation plan can be found at: http://www.energydataweb.com/cpucFiles/pdaDocs/1235/PY2013-2014%20Non-Res%20SO%20Evaluation%20Plan%202015_02_10.pdf

45 California energy efficiency program administrators include PG&E, SCE, SCG, SDG&E, Marin Clean Energy, the Bay Area Regional Energy Network (REN), and the Southern California REN. However, this evaluation only addresses programs under the administration of PG&E, SCE, SCG and SDG&E.

4.2 Weighted NTG Results

Weighted results are presented in this section 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. Since the sample of electric and gas projects was developed based on one common metric, source Btu, net savings results are also being reported by source Btu. Separate reporting by fuel type (electric vs. gas) is not feasible.

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. Weighted NTGRs were calculated for each size stratum for each utility, thereby supporting analysis at the PA-level only.

Note that the Final NTGR values in Table 4-2, Table 4-3, Table 4-4 and Table 4-5 below are based on the removal of 15 projects. This was due to the potential overlap for these sites between the NTG for the project and the Gross ISP (13 projects removed) or Dual Baseline (2 projects removed) determinations in the gross impact analysis.[46] For all PA domains, the improvement in domain level NTGRs from these removals was very slight, on the order of one to two percent.

4.2.1 PG&E Combined Electric and Gas

Table 4-2 below reports NTGRs for all programs implemented by PG&E and represented by the completed surveys.

Table 4-2: Weighted Net-to-Gross Ratios for PG&E – Combined Electric and Gas

Sampling Strata	NTGR
1	0.75
2	0.49
3	0.60
4	0.48
5	0.46
Weighted NTGR	**0.55**
90 Percent CI	0.52 to 0.59
Relative Precision	0.07
n NTGR Completes	51
N Pop Sampling Units	1,126
ER	0.30

[46] A more thorough discussion of baseline designations overturned by the evaluator can be found in the Chapter 5 Project Practices Assessment.

Results by Stratum: Results varied considerably across sample size strata:

- Stratum 1, consisting of the largest projects, had the highest NTGR, that was primarily driven by the results of 3 large oil refinery projects with among the highest NTGRs in the 0.57 to 0.95 range. The other, smaller projects in this stratum had medium-high NTGRs ranging from 0.50 to 0.83.
- Stratum 2 projects had a wide range of results. This size stratum consisted primarily of oil refinery and oil well projects but also included one large greenhouse and 2 data center projects. Results among the oil company projects were mixed, and ranged from lows of 0.03 (2 projects) to 0.81 (2 projects).
- Stratum 3 projects include a mix of oil, food/agriculture, data center and institutional sector projects. NTGRs generally fell in the range of 0.50 and above. Only 3 of the 12 projects evaluated had NTGRs below this level and all were at 0.40 and above.
- Strata 4 and 5 consisted of a wide mix of smaller projects undertaken. Noteworthy results included much lower NTGRs (below 0.40) among sanitation, food/agriculture and data center projects in particular. These results brought the weighted average NTGR at the stratum level down significantly, compared to the other strata.

4.2.2 SCE Electric

In the current evaluation, the SCE weighted NTGR by stratum and across all size projects average 0.57 as shown in Table 4-3 below.

Table 4-3: Weighted Net-to-Gross Ratios for SCE – Electric

Sampling Strata	NTGR
1	0.63
2	0.67
3	0.54
4	0.54
5	0.45
Weighted NTGR	**0.57**
90 Percent CI	0.52 to 0.61
Relative Precision	0.09
n NTGR Completes	46
N Pop Sampling Units	934
ER	0.36

Results by Stratum. The verified NTGR for SCE of 0.57 has increased from the level of 0.49 for PY2010-2012. The largest projects have among the highest NTGRs, for example, the NTGRs for stratum 1 and Stratum 2 projects are 0.63 and 0.67, respectively. NTGRs were somewhat lower for small and medium-sized projects in Strata 3, 4 and 5. Specific drivers are discussed below:

- Stratum 1 and Stratum 2 project results were generally very good. Within Stratum 1, four out of five projects had NTGRs ranging from 0.57 to 0.95. Similarly, Stratum 2 project results for 3 out of 4 projects evaluated ranged from 0.62 to 1.00. Evaluated projects represented a wide range of customer types and installed technologies.
- Smart well project NTGRs were influential in Strata 3, 4 and 5 results. In total, there were 9 such projects with NTGRs averaging 0.50. In addition, Stratum 3 included two statewide university partnership projects exhibiting very high program influence with NTGRs of 0.80 and 0.95, respectively.

4.2.3 SDG&E Combined Electric and Gas

The average NTGR for SDG&E's electric and gas projects is similar in magnitude to SCE's, averaging 0.59 across all projects evaluated. Table 4-4 below reports verified NTGRs across all evaluated projects.

Table 4-4: Weighted Net-to-Gross Ratios for SDGE – Combined Electric and Gas

Sampling Strata	NTGR
1	0.77
2	0.57
3	0.68
4	0.50
5	0.46
Weighted NTGR	**0.59**
90 Percent CI	0.55 to 0.64
Relative Precision	0.08
n NTGR Completes	28
N Pop Sampling Units	264
ER	0.28

Results by Stratum. The overall sample size was fairly small (n = 28), thereby limiting the observations that can be made.

- NTGRs of projects in Strata 1 and 2 range from 0.52 to 1.00, demonstrating a medium to strong program influence. However, this conclusion is based on a very small number of projects in each stratum (n= 2).
- The sample includes several statewide university partnership projects distributed across 3 size strata. All of these projects had perfect NTGRs of 1.00.
- Sampled projects implemented through the Energy Savings BID program had mixed results, and generally did not perform as well as in the PY2010-2012 evaluation. Larger projects in Strata 1 and 2 exhibited high NTGRs ranging from 0.65 to 0.70, while smaller projects in Strata 3, 4 and 5 performed poorly with NTGRs below 0.50.
- NTGRs for smaller projects, particularly those in Strata 4 and 5, ranged broadly in values, from a low of 0.18 to a high of 1.00. The average NTGRs for each of these strata represent the mid-point of the range of values seen.

4.2.4 SCG Gas

For SCG gas projects, the weighted NTGR across all programs and projects is 0.66, as shown in Table 4-5 below.

Table 4-5: Weighted Net-to-Gross Ratios for SCG – Gas

Sampling Strata	NTGR
1	0.87
2	0.67
3	0.56
4	0.48
5	0.64
Weighted NTGR	**0.66**
90 Percent CI	0.59 to 0.73
Relative Precision	0.10
n NTGR Completes	21
N Pop Sampling Units	159
ER	0.31

Results by Stratum. There was wide variation in the results by size stratum.

- NTGR results in the three largest size strata (strata 1, 2 and 3) were dominated by large oil refinery projects exhibiting medium high to high levels of program influence. However, each of these strata had a very small number of projects (Strata 1 and 2: two projects each, Stratum 3: three projects). All of these large refinery projects were very program driven, and had high NTGRs. The overall improvement since PY2010-2012 in SCG's average NTGR across all evaluated projects is dominated by these strong refinery project results.
- Among smaller sized projects in Strata 4 and 5, there was wide variation in the results. NTGRs ranged from 0.05 to 0.80 in Stratum 4 and from 0.26 to 0.75 in Stratum 5.

4.2.5 Comparison of 2013 and 2010-12 NTG Results by PA – Combined MMBtu

In 2010-12 sampling and analysis was performed separately for electric and gas projects, where applicable. To provide a comparison, 2010-12 results were weighted by fuel-based MMBtu and are presented in Table 4-6 and Figure 4-1. These exhibits show that 2013 NTG results are statistically significantly higher than the previous cycle at the 90% confidence level.

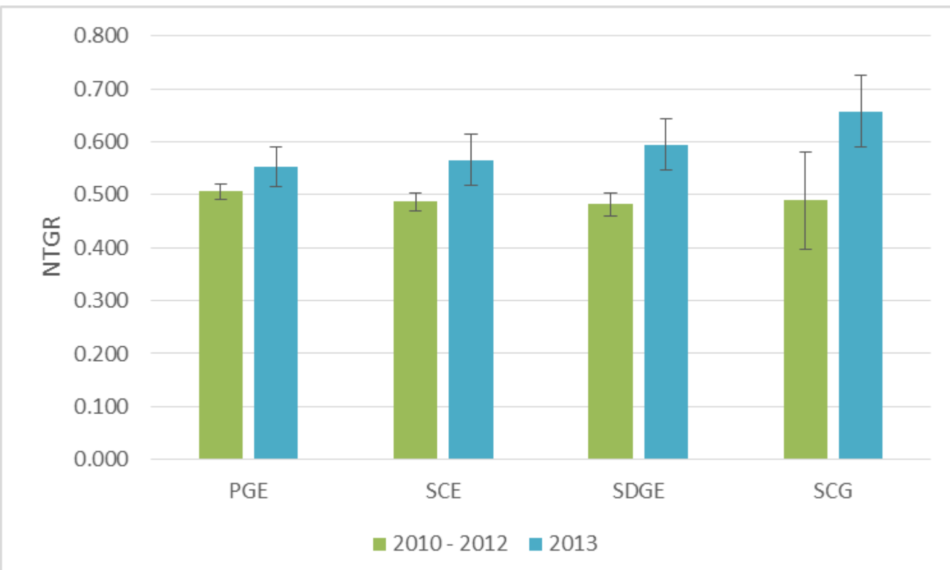
- At the level of PA sampling domain, the final NTGRs range from 0.55 to 0.66, signifying some improvement over the previous cycle.
- **PG&E:** The weighted NTGRs for the PG&E improved by 8 percent compared to PY2010-2012 evaluation results.
- **SCE:** Current cycle results have improved by 16 percent compared to PY2010-2012, based on an NTGR of 0.57 for PY2013 projects versus an NTGR of 0.49 for PY2010-2012.
- **SDGE:** NTGRs for SDG&E's projects have improved by 23 percent compared to PY2010-2012 results, averaging 0.59 across all PY2013 projects evaluated.
- **SCG:** For SCG the weighted NTGR across all projects is 0.66. This represents a 35 percent improvement over the PY2010-2012 average NTGR of 0.49.

Table 4-6: Comparison of 2010-12 and 2013 Weighted MMBtu NTGR Results

Energy Metric	2010-2012 Mean NTGR	2010-2012 90% Confidence Interval	2013 Mean NTGR	2013 90% Confidence Interval
PG&E				
MMBtu*	0.51	0.49 to 0.52	0.55	0.52 to 0.59
SCE				
MMBtu	0.49	0.47 to 0.50	0.57	0.52 to 0.61
SDGE				
MMBtu*	0.48	0.46 to 0.50	0.59	0.55 to 0.64
SCG				
MMBtu	0.49	0.40 to 0.58	0.66	0.59 to 0.73

* The sample for 2010-2012 was *not* designed and selected based on MMBtu.

Figure 4-1: Comparison of 2010-12 and 2013 Weighted MMBtu NTGR Results



4.3 NTG Sensitivity Analysis

NTG ratios were calculated for each of the sample points based on the traditional weighting scheme of equal weights for each of the three scores. A sensitivity analysis of the NTGRs was also conducted, to assess the stability of NTGR results as a function of the weighting scheme. This analysis involved making adjustments to the weights given for each score. In addition to the current weighting scheme of 1/3 to each score, a number of different weighting combinations were analyzed.

Both the weighting schemes and the resulting NTGRs are shown below in Table 4-7. Note that only combination 5 involves reliance on only one of the 3 scores under certain conditions.

Table 4-7: Results of NTG Sensitivity Analysis

NTGR Weighting Scheme	NTGR Result*
1. 33.3% weights to Scores 1, 2 and 3 (current approach)	0.54
2. 50% weight to score 1, 25% to scores 2 and 3	0.54
3. 50% weight to score 2, 25% to scores 1 and 3	0.54
4. 50% weight to score 3, 25% to scores 1 and 2	0.55
5. Use only Score 3 if No-Program Likelihood is 10	0.52

* Based on simple averaging.

These results indicate that the resulting NTGRs are not very sensitive to the weighting scheme used and thus, are highly stable across the different weighting schemes. Only the extreme case (5) when Score 3 was used exclusively if the no-program likelihood of installation was scored a 10 out of 10 showed any variation at all and this was very minor.

In part, this stability exhibited is due to the large number of surveys completed and in part, because of the consistency across the 3 scores. Consistency checking and resolution of inconsistencies is a key part of the NTG survey approach utilized.

4.4 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. Table 4-8 presents the results of this analysis across all projects for which the weighted NTG factors were developed. Results are reported for only the percentage of respondents offering the strongest responses (importance scores of 8, 9 or 10).

The following are general themes and observations across these analyses:

- **As in PY2010-2012, 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.
- Certain factors can contribute to projects with a **lower level of program influence**, and include:

  - Project is in the capital and/or operating budget
  - Measure is installed elsewhere in the company, in places that do not offer rebates
  - Measure automates existing manual processes
  - Measure is associated with environmental compliance
- Other factors that can contribute to **higher program influence** at the project level include:
  - Timing of decision relative to discussions with program staff
  - First-time installation of a measure
  - Energy efficiency is the sole or primary motivation
  - Program rebate is a high percentage of the project cost (>25 percent)

Overall, the presence of factors contributing to higher program influence on projects was more pronounced than those contributing to lower program influence levels.

Table 4-8: NTG Reasons for all PAs

	PG&E	SCE	SCG	SDG&E
Distribution of NTGRs				
High - 0.76 to 1.00	10%	4%	14%	11%
Medium High- 0.51 to 0.75	35%	48%	24%	36%
Medium Low- 0.26 to 0.50	45%	35%	38%	29%
Low - 0.00 to 0.25	10%	13%	24%	25%
NTGR	**0.55**	**0.57**	**0.66**	**0.59**
Key Project Drivers				
Project Maturity				
Project is in the capital and/or operating budget	12%	0%	0%	4%
Project was conceived only after discussions with program staff	22%	24%	24%	32%
Corporate Policy/Practice				
Measure is part of corporate standard practice	45%	65%	43%	61%
Measure is installed elsewhere in company, in places that do not offer rebates	4%	2%	0%	0%
This is a first-time installation of the measure	6%	2%	10%	14%
Company has environmental policy in place	69%	76%	52%	79%
Energy Efficiency A Secondary, not Primary, Benefit				
Measure automates existing manual processes	27%	33%	43%	32%
Energy efficiency is the sole or primary motivation	16%	15%	14%	18%
Measure improves workplace quality	10%	4%	14%	7%
Environmental Compliance				
Measure is associated with environmental compliance (e.g., pollution reduction)	4%	4%	5%	11%
Market Segment				
Measure is installed by a market segment that is ahead of curve on energy efficiency	2%	0%	0%	0%
Measure is installed by national chain/big box firm	14%	7%	14%	18%
Measure is installed by small single location business	24%	30%	29%	18%
Project Cost vs. Rebate				
Rebate is high % of overall project cost (> 25%)	16%	7%	10%	0%
Project Context				
Measure is part of an expansion/remodeling	18%	22%	19%	0%
Measure installed to replace failing equipment	20%	13%	5%	21%

5

Project Practices Assessments

5.1 Introduction

Project Practices Assessments (PPAs) are structured site-specific reviews of Program Administrator (PA)[47] application files and calculations that systematically examine and record the evaluation team's conclusions surrounding PA ex-ante savings development practices. PPAs were completed for each M&V point/measure in the gross impact sample selected for evaluation. The work includes a review of project compliance with CPUC policy and ex-ante review (EAR) guidance, conformance with program rules, use of best practices from industry M&V protocols, and more. Importantly, PPA also supports a comparison between PA and the evaluation team's conclusions. This chapter presents aggregate PPA results across sample points, segmented by PA and application agreement date.

The purpose of the PPA process is to build upon the results of the Low Rigor Assessment (LRA) process that was part of the 10/12 custom impact (WO033) evaluation with the goal of assessing the accuracy and completeness of ex ante parameters recorded and documented in the project files. PPAs are more focused assessments than LRAs and are designed to yield results that can be used to target improvement in PA treatment of important gross impact parameters, methods and procedures that are common across applications. Although PPA assessments generally involve qualitative conclusions of PA work stemming from evaluation M&V efforts, the data generated and the results presented are quantitative.[48]

The PPA process provides impact-oriented findings and feedback to the PAs. The PPA process is conducted on all sampled gross impact points, but analyses and feedback are bifurcated based on applications with a customer agreement date falling in 2013 versus all other applications (pre-2013 and 2013+). This segregation is meant to capture any effects of the policy guidance issued from the 2012 EAR process that might need some lead time to get reflected prospectively in custom project applications (assumed to be approximately one year based on the volume and timing of ex

[47] California energy efficiency program administrators include PG&E, SCE, SCG, SDG&E, Marin Clean Energy, the Bay Area Regional Energy Network (REN), and the Southern California REN. However, this evaluation only addresses programs under the administration of PG&E, SCE, SCG and SDG&E.

[48] By developing results that are presented in a quantitative format, and assuming this work continues for the PY2014 evaluation (as planned) and perhaps for future custom impact evaluations, it will be feasible to use these results as a baseline and measure PA improvement relative to that baseline.

ante reviews).[49] Pre-2013 results serve as an initial baseline against which to measure 2013+ improvement.[50]

5.2 Overview of the Project Practices Assessment

This section briefly describes the assessment process. PPA assessment and reporting feature an examination of the following:

- Project eligibility
- Baseline selection
- Project EUL, RUL, costs and incentives
- Calculation methods, inputs and assumptions

The assessments also directly compare and contrast ex-ante and ex-post conclusions with respect to the above M&V areas. Here the ex-post conclusions represent the evaluator's (and CPUC's) perspective, with differences in ex-ante conclusions representing areas for improvement and agreement representing appropriate ex-ante work that is consistent with CPUC guidance and direction.

These assessments include rating-based examination using the following criteria:

- Comprehensiveness of documentation
- Quality and appropriateness of ex-ante results and conclusions

Special attention was applied to the design of the PPA form and procedure itself so that results can be analyzed objectively to assess conformance with policy guidelines, best practices and program rules.

---

49 http://docs.cpuc.ca.gov/published/FINAL_DECISION/139858.htm. Decision 11-07-030
The EAR process involves an M&V-level of review for PA 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 PA within-program engineering processes and procedures more generally. Importantly, D. 11-07-030 features detailed baseline requirements, including remaining useful life/effective useful life (RUL/EUL) treatment and the need to demonstrate and document all associated early replacement (ER) claims.

50 The evaluation also examined differences between the two periods using both unweighted and sample-weighted FY GRR results, but no consistent patterns emerged. FY GRRs were hypothesized to provide additional evidence of the quality and accuracy of ex-ante engineering-based savings estimates stemming from the PA processes that were examined in this PPA study element, but the results are inconclusive.

The PPA process includes a rating process on a 1 to 5 scale, with 1 representing ex-ante work and conclusions that do not meet basic expectations and 5 representing work and conclusions that consistently exceed expectations. It should be noted that a score of 3 is a desirable score, indicating that the effort meets program expectations. Scores of 4 or 5 are reserved for those applications that went above and beyond typical expectations. Appendix B presents full scoring guidelines used by the evaluation team.

5.3 Project Practices Assessment Results

This report presents unweighted PPA results by PA and aggregated for measure-level records with agreement dates that are pre-2013 versus 2013+. It is noteworthy that resulting sample sizes for the 2013+ period are small for two of the PAs, SDG&E and SCE, at just 9 measures and 14 measures, respectively. These sample sizes are too small to yield meaningful results, and results throughout this section should be reviewed with that key consideration in mind. In total, 240 individual measures from 189 projects were evaluated across all PAs. This section presents and discusses PA and time period-specific results for the most critical aspects of the PPA, especially those identified in Chapter 3 as being primary drivers of discrepancies between ex-ante and ex-post savings (project eligibility, project baselines, project EUL, and calculation methods). Additional results and findings from the PPA analysis can be found in Appendix E (Project Eligibility Considerations, RUL Assessment, Full Cost Assessment, Incremental Cost Assessment, and the Incentive Assessment).

5.3.1 Project Eligibility Ratings

The two primary objectives of this portion of the PPA are to rate the *Quality* of the documentation supporting measure eligibility and then to rate the *Appropriateness* of the PA assessment of eligibility requirements. While every effort was made by the evaluation team to ensure consistency in scoring across measures, it is important to realize that not all custom measures require the same level of documentation. Because of this, all evaluated measures were classified as “routine” or “non-routine.” For the purposes of this analysis, “routine” measures include standard add-on measures (such as VFDs, controls, heat recovery measures, insulation, and etc.), RCx measures, and standard retrofit measures (boilers, chillers, pumps, fans, motors, and etc.). “Non-routine” measures include specialty refrigeration, oil refinery, regenerative thermal oxidizers (RTOs), some non-maintenance IRCx measures, DCV, split pass flow design, natural gas-fired cooking equipment, and any complex system modifications or other uncommon measures. “Routine” measures made up the large majority of measures evaluated (192 of 240) and were not subject to the same documentation expectations as "non-routine" measures. "Routine" measures were typically given quality and appropriateness scores of 3, even with minimal documentation.[51]

Table 5-1 presents eligibility appropriateness and documentation quality scores by PA and application period for all measures evaluated, eligible measures only, and non-routine measures only. Table 5-2 presents the same information for measures determined to be ineligible, along with the reasons cited by the ex-post evaluator for eligibility failure.

Table 5-1 shows that for all measures with eligibility appropriateness scores (233 out of 240), the mean eligibility rating was generally between 2.0 and 2.5 for both the pre-2013 and 2013+ periods (except for - PG&E in the pre period: 2.74; SCE in the post period: 1.92; and SDG&E in the post period 3.0). Mean scores decreased in the 2013+ period relative to the pre-2013 period for PG&E and SCE while they increased for SCG and SDG&E. Unsurprisingly, median eligibility appropriateness ratings were 3.0's across the board due to the large number of "routine" measures. For eligible measures only, appropriateness scores were slightly higher than the overall means (generally between 2.5 and 3.0), with slightly higher scores for 2013+ for three of the four PAs. Very similar results are seen for eligibility documentation rating scores for all measures and for eligible-only measures. However, because of the preponderance of "routine" measures, mean scores for all evaluated points and for eligible-only points are somewhat diluted. The most interesting results in Table 5-1 are those for "non-routine" measures. We see that all mean scores for all PAs (both quality and appropriateness) are between 2.5 and 3.0 for the pre period and all scores are slightly higher at 3.0 for the 2013+ period. It is particularly important that these measures are well documented and appropriately assessed for eligibility.

Table 5-2 shows that 31 measures were found to be ineligible in the pre-2013 period and seven measures in the 2013+ period across the four PAs. Mean appropriateness scores for ineligible measures range from 1.29 to 2.0 in the pre-2013 period and from 1.0 to 2.0 in the post period. Similarly, eligibility quality scores for these measures range from 1.43 to 2.75 in the pre-period and 1 to 2.25 in the post-period. It may be expected that all ineligible measures would receive quality and appropriateness scores of 1. However, quality scores may be greater than 1 if, for example, the PA provided documentation that was satisfactory, but the PA assessment of eligibility based on that documentation was inappropriate. Appropriateness scores may be greater than 1 for an ineligible project due to several factors. For instance, there were cases where applicable CPUC guidance was issued only slightly in advance of the customer agreement, likely leaving inadequate time for the PA to address all relevant and active applications.

[51] As mentioned in Chapter 3, there are a number of cases where ISP or code baseline determination rendered a project ineligible. In these cases where project eligibility and baseline are directly linked, evaluation engineers gave eligibility quality and appropriateness scores of 1 or 2 if documentation was insufficient to establish above code/ISP performance, even for "routine measures."

Table 5-2 also presents the ex-post conclusions for why measures were determined to be ineligible. While a variety of reasons for ineligibility were cited, the majority of ineligible projects were due to CPUC guidance, requirement that measures exceed code/ISP baseline, and previous EAR guidance. Additional detail on eligibility considerations can be found in Appendix E.

Table 5-1: PA Eligibility Ratings by Customer Agreement Date

Parameter Examined	PA Eligibility Ratings (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)							
	Pre-2013 Customer Agreement Date				2013+ Customer Agreement Date			
	PG&E	SCE	SCG	SDG&E	PG&E	SCE	SCG	SDG&E
Number of Measures Assessed	42	50	22	56	27	14	20	9
Assessment of PA Eligibility Appropriateness Rating								
Number of Measures w/ Eligibility Appropriateness Ratings (n)	42	50	22	56	27	14	20	9
Mean Eligibility Appropriateness Rating	2.74	2.47	2.07	2.11	2.17	1.92	2.34	3.00
Median Eligibility Appropriateness Rating	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Number of ELIGIBLE Measures w/ Eligibility Appropriateness Ratings (n)	38	43	16	53	25	13	16	9
Mean Eligibility Appropriateness Rating for ELIGIBLE Measures	2.74	2.79	2.81	2.55	2.84	2.85	2.44	3.00
Median Eligibility Appropriateness Rating for ELIGIBLE Measures	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Number of NON-ROUTINE Measures w/ Eligibility Appropriateness Ratings (n)	12	9	10	3	2	2	5	5
Mean Eligibility Appropriateness Rating for NON-ROUTINE Measures	3.00	2.89	2.50	3.00	3.00	3.00	3.00	3.00
Median Eligibility Appropriateness Rating for NON-ROUTINE Measures	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Assessment of PA Quality of Eligibility Documentation Rating								
Number of Measures w/ Quality of Eligibility Documentation Ratings (n)	42	50	22	56	27	14	20	9
Mean Quality of Eligibility Documentation Rating	2.36	2.04	2.18	2.15	2.19	1.92	2.25	3.00
Median Quality of Eligibility Documentation Rating	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Number of ELIGIBLE Measures w/ Quality of Eligibility Documentation Ratings (n)	38	43	16	53	25	13	16	9
Mean Quality of Eligibility Documentation Rating for ELIGIBLE Measures	2.71	2.79	2.69	2.64	2.88	2.85	2.50	3.00
Median Quality of Eligibility Documentation Rating for ELIGIBLE Measures	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Number of NON-ROUTINE Measures w/ Quality of Eligibility Doc. Ratings (n)	12	9	10	3	2	2	5	5
Mean Quality of Eligibility Documentation Rating for NON-ROUTINE Measures	2.67	2.67	2.60	3.00	3.00	3.00	3.00	3.00
Median Quality of Eligibility Documentation Rating for NON-ROUTINE Measures	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0

Table 5-2: PA Eligibility Ratings for INELIGIBLE MEASURES by Customer Agreement Date, and Ex-Post M&V Conclusions Why Measures are INELIGIBLE

Parameter Examined	PA Eligibility Treatment (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)							
	Pre-2013 Customer Agreement Date				2013+ Customer Agreement Date			
	PG&E	SCE	SCG	SDG&E	PG&E	SCE	SCG	SDG&E
Number of Measures Evaluated	42	50	22	56	27	14	20	9
Number of INELIGIBLE Measures	4	7	7	13	2	1	4	0
Percent of All Measures Found to be INELIGIBLE	10%	14%	32%	23%	7%	7%	20%	0%
Mean Appropriateness of Eligibility Score	2.00	1.29	1.71	1.54	1.50	1.00	2.00	--
Median Appropriateness of Eligibility Score	2.0	1.0	2.0	1.0	1.5	1.0	2.0	--
Mean Quality of Documentation Score	2.75	2.14	1.43	1.46	1.50	1.00	2.25	--
Median Quality of Documentation Score	3.0	3.0	1.0	1.0	1.5	1.0	2.5	--
Number of NON-ROUTINE Measures Found to be INELIGIBLE	2	1	3	--	--	--	2	--
Evaluation Conclusions Why Measures are INELIGIBLE								
Conclusions								
Program rules	1	--	--	--	--	--	--	--
Normal maintenance	--	1	4	--	--	--	1	--
Operating practice change	--	--	--	1	--	--	1	--
CPUC decisions	--	1	--	2	--	--	1	--
CPUC guidance	--	3	--	2	1	--	1	--
Requirement that measures exceed code / ISP baseline	2	6	4	5	2	1	1	--
Previous EAR guidance	1	3	--	2	1	1	--	--
Previous evaluation findings	1	1	--	2	--	--	--	--
Project boundary condition	--	--	--	--	--	--	--	--
EE Policy Manual	--	--	--	--	--	--	--	--
Multiple PA fuels (includes cogeneration and fuel switching)	--	--	--	--	--	--	--	--
Three prong test	--	--	--	--	--	--	--	--
Non-PA fuels and ancillary impacts (i.e., cogen, refinery gas, WHR, etc.)	--	--	--	--	--	--	--	--
Regressive Baseline Rule	1	--	--	--	--	--	--	--
Savings not demonstrated	--	--	--	--	--	--	2	--
Redundant equipment	--	--	--	1	--	--	--	--
Other	1	--	--	--	--	--	--	--

5.3.2 Project Types

Table 5-3 through Table 5-6 present PA-specific results detailing ex-ante versus ex-post project type designations. The shaded cells along the diagonal indicate the number of measures that showed agreement between the PA and ex-post evaluation. Values in the non-shaded cells are measures where the project type was reassigned by the evaluator. In general, the analysis shows relatively good agreement between PA and evaluator specified project types and three of the four PAs showed improvement in project type designation between the pre-2013 and 2013+ periods:

- PG&E: pre-2013 79 percent, 2013+ 85 percent;
- SCE: pre-2013 70 percent, 2013+ 64 percent;
- SCG: pre-2013 64 percent, 2013+ 75 percent;
- SDG&E: pre-2013 73 percent, 2013+ 100 percent.

While perfect agreement between PA and evaluator specified project types reduces the likelihood of evaluated savings becoming vastly different from ex ante savings, it is important to realize that not all project type reassignments have an impact on evaluated FY or LC savings. For example, an IOU project classified as an add-on measure may be reclassified as system optimization without any impact on first year savings. An ER project evaluated as an ROB project or vice versa can have significantly impact the FY and LC GRRs.

Across all PAs, the most commonly overturned project types were early replacement and replace-on-burnout measures. In the pre-2013 period, PA assigned early replacement measures were overturned 48 percent of the time (to NR, ROB, and add-on) and ROB measures were overturned in 53 percent of cases (to ER, NR, Add-on, and SysOp). In the 2013+ period, ER and ROB measures were overturned in 18 percent and 33 percent of cases, respectively. These results are somewhat mixed, however, across the PAs. For example, SCG replace-on-burnout measures were in complete agreement with the evaluation. The reader is referred to the tables for examination of individual PA results.

Table 5-3: PA vs. Evaluation Specified Project Type by Customer Agreement Date – PG&E

			PG&E-Specified Project Type						
			Early Replacement	Natural Replacement	Replace-on-Burnout	Capacity Expansion	New Construction	Add-on Measure	System Optimization
Pre-2013 Customer Agreement Date									
Number of measures evaluated (n)			42						
Frequency of PG&E-Specified Measure Type (n)			1	6	0	0	9	14	12
Evaluation-Specified Project Type	Frequency of Measure-Level Observations (n)								
	Early replacement	1	0	0	0	0	0	1	0
	Natural replacement	6	0	5	0	0	0	1	0
	Replace-on-burnout	1	1	0	0	0	0	0	0
	Capacity expansion	4	0	1	0	0	3	0	0
	New construction	6	0	0	0	0	6	0	0
	Add-on measure	12	0	0	0	0	0	11	1
	System optimization	12	0	0	0	0	0	1	11
2013+ Customer Agreement Date									
Number of measures evaluated (n)			27						
Frequency of PG&E-Specified Measure Type (n)			2	2	2	2	4	9	6
Evaluation-Specified Project Type	Frequency of Measure-Level Observations (n)								
	Early replacement	1	1	0	0	0	0	0	0
	Natural replacement	3	0	2	1	0	0	0	0
	Replace-on-burnout	3	1	0	1	0	0	1	0
	Capacity expansion	2	0	0	0	2	0	0	0
	New construction	4	0	0	0	0	4	0	0
	Add-on measure	7	0	0	0	0	0	7	0
	System optimization	7	0	0	0	0	0	1	6

Table 5-4: PA vs. Evaluation Specified Project Type by Customer Agreement Date – SCE

			SCE-Specified Project Type						
			Early Replacement	Natural Replacement	Replace-on-Burnout	Capacity Expansion	New Construction	Add-on Measure	System Optimization
Pre-2013 Customer Agreement Date									
Number of measures evaluated (n)			50						
Frequency of SCE-Specified Measure Type (n)			6	0	12	1	5	18	8
Evaluation-Specified Project Type	Frequency of Measure-Level Observations (n)								
	Early replacement	6	5	0	0	0	0	1	0
	Natural replacement	3	0	0	3	0	0	0	0
	Replace-on-burnout	6	1	0	5	0	0	0	0
	Capacity expansion	3	0	0	0	1	2	0	0
	New construction	7	0	0	0	0	3	4	0
	Add-on measure	16	0	0	3	0	0	13	0
	System optimization	9	0	0	1	0	0	0	8
2013+ Customer Agreement Date									
Number of measures evaluated (n)			14						
Frequency of SCE-Specified Measure Type (n)			2	1	1	0	1	6	3
Evaluation-Specified Project Type	Frequency of Measure-Level Observations (n)								
	Early replacement	3	2	0	0	0	0	1	0
	Natural replacement	1	0	1	0	0	0	0	0
	Replace-on-burnout	0	0	0	0	0	0	0	0
	Capacity expansion	2	0	0	0	0	1	0	1
	New construction	1	0	0	0	0	0	1	0
	Add-on measure	5	0	0	1	0	0	4	0
	System optimization	2	0	0	0	0	0	0	2

Table 5-5: PA vs. Evaluation Specified Project Type by Customer Agreement Date – SCG

			SCG-Specified Project Type						
			Early Replacement	Natural Replacement	Replace-on-Burnout	Capacity Expansion	New Construction	Add-on Measure	System Optimization
Pre-2013 Customer Agreement Date									
Number of measures evaluated (n)			22						
Frequency of SCG-Specified Measure Type (n)			4	0	2	1	4	7	4
Evaluation-Specified Project Type	Frequency of Measure-Level Observations (n)								
	Early replacement	0	0	0	0	0	0	0	0
	Natural replacement	4	2	0	0	0	0	0	2
	Replace-on-burnout	4	2	0	2	0	0	0	0
	Capacity expansion	2	0	0	0	1	1	0	0
	New construction	3	0	0	0	0	3	0	0
	Add-on measure	8	0	0	0	0	0	7	1
	System optimization	1	0	0	0	0	0	0	1
2013+ Customer Agreement Date									
Number of measures evaluated (n)			20						
Frequency of SCG-Specified Measure Type (n)			2	1	3	2	2	8	2
Evaluation-Specified Project Type	Frequency of Measure-Level Observations (n)								
	Early replacement	1	1	0	0	0	0	0	0
	Natural replacement	1	0	1	0	0	0	0	0
	Replace-on-burnout	7	1	0	3	0	0	1	2
	Capacity expansion	2	0	0	0	2	0	0	0
	New construction	2	0	0	0	0	2	0	0
	Add-on measure	6	0	0	0	0	0	6	0
	System optimization	1	0	0	0	0	0	1	0

Table 5-6: PA vs. Evaluation Specified Project Type by Customer Agreement Date – SDG&E

			SDG&E-Specified Project Type						
			Early Replacement	**Natural Replacement**	**Replace-on-Burnout**	**Capacity Expansion**	**New Construction**	**Add-on Measure**	**System Optimization**
Pre-2013 Customer Agreement Date									
Number of measures evaluated (n)			56						
Frequency of SDG&E-Specified Measure Type (n)			10	3	5	0	3	25	10
Evaluation-Specified Project Type	Frequency of Measure-Level Observations (n)								
	Early replacement	10	6	2	2	0	0	0	0
	Natural replacement	1	0	1	0	0	0	0	0
	Replace-on-burnout	4	1	0	2	0	0	1	0
	Capacity expansion	0	0	0	0	0	0	0	0
	New construction	2	0	0	0	0	2	0	0
	Add-on measure	23	3	0	1	0	1	19	0
	System optimization	15	0	0	0	0	0	5	10
2013+ Customer Agreement Date									
Number of measures evaluated (n)			9						
Frequency of SDG&E-Specified Measure Type (n)			5	0	0	0	0	2	2
Evaluation-Specified Project Type	Frequency of Measure-Level Observations (n)								
	Early replacement	5	5	0	0	0	0	0	0
	Natural replacement	0	0	0	0	0	0	0	0
	Replace-on-burnout	0	0	0	0	0	0	0	0
	Capacity expansion	0	0	0	0	0	0	0	0
	New construction	0	0	0	0	0	0	0	0
	Add-on measure	2	0	0	0	0	0	2	0
	System optimization	2	0	0	0	0	0	0	2

5.3.3 Project Type Documentation Ratings

Table 5-7 presents a summary of the quality of the PA documentation that was used in establishing project baseline types (ER, ROB, etc.) and project baselines (Title 24, ISP, etc.). In the pre-2013 period, quality of documentation scores for all PAs were between 2.0 and 2.5, indicating that on average, the PAs performed "minimal effort to meet expectations." In the 2013+ period, mean quality scores came closer to "effort meets expectations" for PG&E and SCE (2.74 and 2.50, respectively) while scores remained low (2.1 and 1.78) for SCG and SDG&E. In the pre-2013 period, the mean documentation quality scores were skewed low, in part, due to a non-trivial number of measures having no clear baseline documentation (ranging from 14 percent of measures for SCE to 32 percent of measures for SDG&E). Instances of no baseline documentation dropped to below 10 percent for PG&E and SCE in the 2013+ period, while SCG and SDG&E had no documentation for 30 percent and 66 percent of measures.

Table 5-7 also documents the distribution of supporting sources used to establish project types and baselines by the PA and by the evaluator for the pre-2013 and 2013+ periods. For any given measure, multiple sources may have been utilized in establishing project types and baselines (i.e. percentages may sum to greater than 100 percent). Cells highlighted in grey indicate instances where the evaluator consulted a given source 10 percent or more frequently than the PA. These are only "Quality of Documentation" scores, so no direct link can be established between instances of inadequate documentation quality and overturned project types or project baselines established in Table 5-3 through Table 5-6 (i.e. the PA may have provided appropriate documentation but then made the wrong conclusion on project type / baseline or vice versa). However, some conclusions may still be drawn. The preponderance of inadequate PA documentation surrounding the age and condition of existing equipment, EULs/RULs, capabilities of existing / baseline equipment to meet service capabilities, and available efficiency levels likely contributed to the discrepancies between the PAs and evaluators regarding ER, NR, and ROB project types. Similarly, the evaluators more often examined "normal facility practices" which helped to identify instances of regressive baselines and subsequent project ineligibility. Finally, for some PAs, evaluators more often reviewed ISP and applicable codes and standards, which were commonly reasons for overturned project baselines (see section 5.3.4). It should be noted that frequency of citing at least one source is greatest for PG&E and SCE, but then drops for SCG and SDG&E.

Table 5-7: Project Baseline Type Documentation Quality by PA and Customer Agreement Date

Parameter Examined	Quality of PA Baseline Documentation (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)															
	Pre-2013 Customer Agreement Date								2013+ Customer Agreement Date							
	PG&E		SCE		SCG		SDG&E		PG&E		SCE		SCG		SDG&E	
Number of Measures Evaluated	42		50		22		56		27		14		20		9	
Mean Quality of Documentation Rating	2.41		2.34		2.32		2.09		2.74		2.50		2.10		1.78	
Median Quality of Documentation Rating	3.0		3.0		2.0		2.0		3.0		3.0		2.0		2.0	
Comparison of PA Application- and Evaluation-Based Baseline Documentation Sources																
Sources	**Ex-Ante**	**Ex-Post**	**Ex-Ante**	**Ex-Post**	**Ex-Ante**	**Ex-Post**	**Ex-Ante**	**Ex-Post**	**Ex-Ante**	**Ex-Post**	**Ex-Ante**	**Ex-Post**	**Ex-Ante**	**Ex-Post**	**Ex-Ante**	**Ex-Post**
Not documented	17%	--	14%	--	18%	--	32%	--	7%	--	7%	--	30%	--	66%	--
Age of existing equipment	21%	31%	10%	26%	--	14%	7%	34%	7%	26%	29%	29%	10%	15%	--	11%
EUL of equipment	14%	19%	14%	20%	5%	23%	2%	13%	--	33%	29%	29%	15%	20%	--	--
Assessment of RUL	7%	14%	6%	16%	--	18%	4%	18%	11%	26%	14%	36%	--	5%	--	--
Condition of existing equipment	31%	48%	18%	48%	32%	64%	23%	59%	26%	41%	14%	29%	45%	60%	--	100%
Capability of existing equipment to meet service requirements	17%	45%	26%	50%	14%	36%	7%	43%	11%	26%	21%	64%	15%	30%	--	33%
Capability of baseline equipment to meet service requirements	5%	17%	8%	20%	5%	18%	18%	14%	7%	19%	14%	14%	5%	40%	--	--
Equipment replacement schedule	5%	5%	--	4%	--	--	--	--	--	15%	--	--	--	--	--	--
Efficiency level(s) available	40%	55%	18%	38%	18%	36%	2%	14%	44%	52%	21%	50%	20%	35%	--	--
Maintenance records examined	2%	7%	2%	2%	14%	14%	--	2%	--	--	--	--	--	10%	--	22%
Other equipment choices considered	2%	7%	--	--	5%	9%	--	2%	4%	11%	--	7%	--	--	--	--
Other equipment choices documented from vendor or designer	--	2%	--	2%	--	--	2%	2%	7%	7%	--	--	--	--	--	--
Normal facility practices examined (regressive baseline)	21%	55%	36%	48%	32%	41%	23%	25%	19%	52%	36%	71%	30%	45%	33%	33%
Standard industry practices examined	17%	21%	16%	26%	18%	32%	--	5%	--	11%	7%	14%	5%	5%	--	--
Standard industry practices researched	7%	7%	2%	2%	--	18%	2%	2%	4%	4%	--	14%	--	5%	--	--
Applicable codes or regulations examined	12%	19%	6%	8%	5%	5%	4%	13%	19%	30%	--	--	5%	15%	--	--
Written facility plans if "no program"	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	11%
Stated facility plans if "no program"	--	--	--	2%	--	--	--	--	--	7%	--	7%	--	--	--	--
NTG project screening interview completed	--	2%		2%	--	--	--	--	--	--	--	--	--	--	--	--
Miscellaneous other	7%	21%	--	4%	5%	--	5%	2%	7%	7%	7%	--	--	--	--	--

5.3.4 Project Baselines

Table 5-8 through Table 5-11 present an overview of the PA versus ex-post project baseline designations.[52] As with project type, the shaded cells along the diagonal indicate the number of measures that showed agreement between the PA and ex-post evaluation, and values in the non-shaded cells are measures where the project baseline was reassigned by the evaluator. Again, the analysis shows improvement in project baseline designation between the pre-2013 and 2013+ periods, but there is room for significant improvement overall:

- PG&E: pre-2013 70 percent, 2013+ 81 percent;
- SCE: pre-2013 77 percent, 2013+ 79 percent;
- SCG: pre-2013 68 percent, 2013+ 70 percent;
- SDG&E: pre-2013 70 percent, 2013+ 100 percent.

For all PAs, project baselines were most often overturned by the evaluator in cases where the PA specified existing equipment as the baseline. In the cases where the in-situ baseline was overturned, the proper baseline was most often established through ISP or state / federal codes, especially in the pre-2013 period. It is typically the case that the existing equipment is less efficient that the applicable ISP or code and overturning the baseline results in a negative impact on savings (recall that this was observed as an influential discrepancy factor in Chapter 3).

[52] Note that for some PAs, the frequency of PA or evaluator specified baseline types exceed the number of measures evaluated. This is due to some individual tracking records that consist of two or more measures with different applicable baselines.

Table 5-8: PA vs. Evaluation Specified Project Baseline by Customer Agreement Date – PG&E

			PG&E-Specified Project Baseline								
			Existing Equipment	Title 24	Industry Standard Practice	Title 20	Customer / Facility Standard Practice	Local AQMD/ Other Code	Federal Regulations	Misc. / Other[53]	None Specified
Pre-2013 Customer Agreement Date											
Number of measures evaluated (n)			42								
Frequency of PG&E-Specified Measure Baseline (n)			32	0	4	3	0	0	0	4	0
Evaluation-Specified Project Baseline	Frequency of Measure-Level Observations (n)										
	Existing equipment	25	24	0	0	0	0	0	0	1	0
	Title 24	1	1	0	0	0	0	0	0	0	0
	Industry standard practice	7	3	0	4	0	0	0	0	0	0
	Title 20	1	1	0	0	0	0	0	0	0	0
	Customer/facility standard practice	0	0	0	0	0	0	0	0	0	0
	Local AQMD/other code	0	0	0	0	0	0	0	0	0	0
	Federal regulations	2	1	0	0	0	0	0	0	1	0
	Miscellaneous / other	7	2	0	0	3	0	0	0	2	0
2013+ Customer Agreement Date											
Number of measures evaluated (n)			27								
Frequency of PG&E-Specified Measure Type (n)			18	3	2	0	1	1	0	0	2
Evaluation-Specified Project Baseline	Frequency of Measure-Level Observations (n)										
	Existing equipment	16	16	0	0	0	0	0	0	0	0
	Title 24	4	0	3	0	0	0	0	0	0	1
	Industry standard practice	5	2	0	2	0	0	0	0	0	1
	Title 20	0	0	0	0	0	0	0	0	0	0
	Customer/facility standard practice	0	0	0	0	0	0	0	0	0	0
	Local AQMD/other code	1	0	0	0	0	0	1	0	0	0
	Federal regulations	0	0	0	0	0	0	0	0	0	0
	Miscellaneous / other	1	0	0	0	0	1	0	0	0	0

[53] Miscellaneous / Other project baseline designations may include CPUC staff directives, dedicated baseline studies or workpapers, or unique project-specific combinations of two or more baseline considerations, among others.

Table 5-9: PA vs. Evaluation Specified Project Baseline by Customer Agreement Date – SCE

			SCE-Specified Project Baseline							
			Existing Equipment	Title 24	Industry Standard Practice	Title 20	Customer / Facility Standard Practice	Local AQMD/ Other Code	Federal Regulations	Misc. / Other
Pre-2013 Customer Agreement Date										
Number of measures evaluated (n)			50							
Frequency of SCE-Specified Measure Baseline (n)			37	6	6	0	1	0	0	2
Evaluation-Specified Project Baseline	Frequency of Measure-Level Observations									
	Existing equipment	29	28	0	0	0	0	0	0	1
	Title 24	7	0	6	1	0	0	0	0	0
	Industry standard practice	12	6	0	5	0	1	0	0	0
	Title 20	0	0	0	0	0	0	0	0	0
	Customer/facility standard practice	1	1	0	0	0	0	0	0	0
	Local AQMD/other code	0	0	0	0	0	0	0	0	0
	Federal regulations	0	0	0	0	0	0	0	0	0
	Miscellaneous / other	3	2	0	0	0	0	0	0	1
2013+ Customer Agreement Date										
Number of measures evaluated (n)			14							
Frequency of SCE-Specified Measure Type (n)			11	0	3	0	0	0	0	0
Evaluation-Specified Project Baseline	Frequency of Measure-Level Observations									
	Existing equipment	8	8	0	0	0	0	0	0	0
	Title 24	0	0	0	0	0	0	0	0	0
	Industry standard practice	4	1	0	3	0	0	0	0	0
	Title 20	0	0	0	0	0	0	0	0	0
	Customer/facility standard practice	0	0	0	0	0	0	0	0	0
	Local AQMD/other code	0	0	0	0	0	0	0	0	0
	Federal regulations	0	0	0	0	0	0	0	0	0
	Miscellaneous / other	2	2	0	0	0	0	0	0	0

Table 5-10: PA vs. Evaluation Specified Project Baseline by Customer Agreement Date – SCG

			SCG-Specified Project Baseline								
			Existing Equipment	**Title 24**	**Industry Standard Practice**	**Title 20**	**Customer / Facility Standard Practice**	**Local AQMD/ Other Code**	**Federal Regulations**	**Misc. / Other**	**None Specified**
Pre-2013 Customer Agreement Date											
Number of measures evaluated (n)			22								
Frequency of SCG-Specified Measure Baseline (n)			15	0	5	0	0	0	0	0	2
Evaluation-Specified Project Baseline	Frequency of Measure-Level Observations (n)										
	Existing equipment	10	10	0	0	0	0	0	0	0	0
	Title 24	0	0	0	0	0	0	0	0	0	0
	Industry standard practice	10	3	0	5	0	0	0	0	0	2
	Title 20	0	0	0	0	0	0	0	0	0	0
	Customer/facility standard practice	0	0	0	0	0	0	0	0	0	0
	Local AQMD/other code	0	0	0	0	0	0	0	0	0	0
	Federal regulations	0	0	0	0	0	0	0	0	0	0
	Miscellaneous / other	2	2	0	0	0	0	0	0	0	0
2013+ Customer Agreement Date											
Number of measures evaluated (n)			20								
Frequency of SCG-Specified Measure Type (n)			14	0	1	0	3	0	0	2	0
Evaluation-Specified Project Baseline	Frequency of Measure-Level Observations (n)										
	Existing equipment	12	12	0	0	0	0	0	0	0	0
	Title 24	1	1	0	0	0	0	0	0	0	0
	Industry standard practice	2	0	0	1	0	0	0	0	1	0
	Title 20	0	0	0	0	0	0	0	0	0	0
	Customer/facility standard practice	0	0	0	0	0	0	0	0	0	0
	Local AQMD/other code	1	0	0	0	0	1	0	0	0	0
	Federal regulations	0	0	0	0	0	0	0	0	0	0
	Miscellaneous / other	4	1	0	0	0	2	0	0	1	0

Table 5-11: PA vs. Evaluation Specified Project Baseline by Customer Agreement Date – SDG&E

			SDG&E-Specified Project Baseline							
			Existing Equipment	Title 24	Industry Standard Practice	Title 20	Customer / Facility Standard Practice	Local AQMD/ Other Code	Federal Regulations	Misc. / Other
Pre-2013 Customer Agreement Date										
Number of measures evaluated (n)			56							
Frequency of SDG&E-Specified Measure Baseline (n)			51	3	0	0	0	0	0	3
Evaluation-Specified Project Baseline	Frequency of Measure-Level Observations									
	Existing equipment	35	35	0	0	0	0	0	0	0
	Title 24	8	5	3	0	0	0	0	0	0
	Industry standard practice	2	2	0	0	0	0	0	0	0
	Title 20	0	0	0	0	0	0	0	0	0
	Customer/facility standard practice	0	0	0	0	0	0	0	0	0
	Local AQMD/other code	0	0	0	0	0	0	0	0	0
	Federal regulations	2	1	0	0	0	0	0	0	1
	Miscellaneous / other	10	8	0	0	0	0	0	0	2
2013+ Customer Agreement Date										
Number of measures evaluated (n)			9							
Frequency of SDG&E-Specified Measure Type (n)			9	0	0	0	0	0	0	0
Evaluation-Specified Project Baseline	Frequency of Measure-Level Observations									
	Existing equipment	9	9	0	0	0	0	0	0	0
	Title 24	0	0	0	0	0	0	0	0	0
	Industry standard practice	0	0	0	0	0	0	0	0	0
	Title 20	0	0	0	0	0	0	0	0	0
	Customer/facility standard practice	0	0	0	0	0	0	0	0	0
	Local AQMD/other code	0	0	0	0	0	0	0	0	0
	Federal regulations	0	0	0	0	0	0	0	0	0
	Miscellaneous / other	0	0	0	0	0	0	0	0	0

5.3.5 Project Baseline Ratings

Table 5-12 includes scores for baseline appropriateness and baseline description. Based on review of the relevant documentation, baseline appropriateness scores reflect whether or not the PA correctly identified the project baseline (existing equipment, Title 24, etc.). For measures with dual baseline considerations, this score also includes whether the second baseline was accurately specified and included in the lifecycle savings calculations in the project documentation. Low baseline appropriateness scores generally correspond to overturned baselines demonstrated in Table 5-8 through Table 5-11. Baseline description scores capture the extent to which PAs described the actual baseline equipment and the efficiency rating of that equipment. Again, for dual baselines, both baselines must be accurately described, including descriptions of the EUL and RUL periods. These descriptions and ratings are critical for accurate savings calculations.

Mean scores for both appropriateness and baseline description showed minimal changes between the pre-2013 and 2013+ periods. All mean scores for these two PPA criteria were between 2.2 and 2.81. Table 5-12 also lists the frequency with which measures received appropriateness ratings of 1 or 2. SCG and SDG&E receive scores of 1 or 2 most frequently relative to total sample size, explaining to a large degree the lower overall ratings for those two PAs. Similarly, Table 5-12 shows the frequency of baseline description ratings of 1 or 2. Again, relative to total sample size, SCG and SDG&E are more likely to receive scores of 1 or 2 for this metric.

Table 5-12: Project Baseline Ratings by PA and Customer Agreement Date

Parameter Examined	PA Baseline Ratings (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)							
	Pre-2013 Customer Agreement Date				2013+ Customer Agreement Date			
	PG&E	SCE	SCG	SDG&E	PG&E	SCE	SCG	SDG&E
Number of Measures Assessed	42	50	22	56	27	14	20	9
Percent of Measures with PA Baseline Documented in the Project Application File	100%	100%	91%	100%	93%	100%	100%	100%
Assessment of PA Baseline Appropriateness Rating								
Number of Measures with Baseline Appropriateness Ratings (n)	42	50	22	56	27	14	20	9
Mean Baseline Appropriateness Rating	2.62	2.72	2.23	2.39	2.78	2.57	2.20	2.22
Median Baseline Appropriateness Rating	3.0	3.0	2.5	3.0	3.0	3.0	2.5	2.0
Number of Measures with Baseline Appropriateness Ratings of 1 or 2 (n)	11	10	11	27	7	4	9	7
Assessment of PA Baseline Description Rating								
Number of Measures with Baseline Description Ratings (n)	42	50	22	56	27	14	20	9
Mean Baseline Description Rating	2.52	2.80	2.55	2.52	2.81	2.64	2.50	2.22
Median Baseline Description Rating	3.0	3.0	3.0	3.0	3.0	3.0	3.0	2.0
Number of Measures with Baseline Description Ratings of 1 or 2 (n)	13	9	9	22	5	2	7	7

5.3.6 EUL Assessment

Table 5-13 provides a comparison of the EUL values that were documented in the PA tracking data, project application, and the ex-post site reports and provides a summary of the EUL sources. Notably, EUL was poorly documented in the project application files compared to the tracking data where EUL was 100 percent documented. Three of the four PAs had declining rates of documentation from the pre-2013 to the 2013+ period and the highest rate of documentation was 64 percent, which indicates that all PAs are providing insufficient EUL data in project files.

- PG&E's EUL documentation declined significantly from the pre-2013 period to the 2013+ period; 33 percent of their sites had EUL in the project application files in the former period and only 4 percent of sites had EUL documented in the project application files in the latter period (an 89 percent decline in documentation).
- Southern California Edison improved their EUL documentation from 46 percent of sites in the pre-2013 period to 64 percent of sites in the 2013+ period, which is a 40 percent increase.
- SoCal Gas's rate of EUL documentation was similar for the pre-2013 and 2013+ periods at 41 percent to 40 percent, respectively.
- SDG&E's documentation of EUL in the projects' application files declined from 46 percent to 11 percent from the pre-2013 to the 2013+ period, which is a 76 percent decrease, although the sample size is small for 2013+.

Mean scores received by the PAs for EUL documentation were between 1.2 and 2.2, with a median of 1 for most PA/time period observations. SCE had median scores of 2 for both time periods and SCG had a median score of 3 for the 2013+ time period.

Of the 91 measures that had EUL documented in the project application files, 43 (47 percent) were sourced from DEER. No source was provided for 36 of the EULs documented in the project applications. Similarly, DEER was the primary source of evaluation-sourced EULs; 133 out of 240 evaluation EULs were sourced from DEER (55 percent).

Across all PAs the mean evaluation-sourced EULs were lower than the mean tracking and project application EULs. For the 46 percent of measures that had different PA tracking and evaluation-sourced EULs (a total of 111 out of 240 evaluated measures across all PAs and time periods), the average differences between the EULs are smaller in the 2013+ period than the pre-2013 period for all PAs except for SDG&E. The mean differences between the evaluation-sourced and tracking EULs for each PA and each application period are as follows:

- PGE - pre-2013: -5.20, 2013+: -2.46;
- SCE - pre-2013: -3.17; 2013+: -2.17;
- SCG - pre-2013: -7.71; 2013+: -0.88;

- SDG&E - pre-2013: -6.14; 2013+: -7.20.

When the EUL differences noted above are combined across all PAs and time periods, the simple average evaluation EUL is approximately 4.5 years less than the average EUL from tracking. If all measures are included (not just those showing differences), the simple average evaluation EUL is approximately 2 years less than the average EUL from tracking. As noted in Chapter 3 LC GRR results were substantially lower than FY GRR results and this EUL difference was *the key factor* driving down the LC GRR results.

The primary reasons for differences among the evaluation and the tracking EULs were noted as being due to different interpretations or applications of DEER (32 percent). Two other common reasons given for differing EULs were that CPUC policy (23 percent)[54] and RCx program rules (17 percent) defined EUL values for some measures that differed from the tracking data EUL values. The large discrepancy in PA EUL values relative to evaluation results is likely due to the PA practice of assigning EULs at the time of making savings claims (in tracking) as compared to including documented EUL in the project file upon project approval.

[54] Specifically, where CPUC guidance stipulates that the EUL may be limited to the RUL of the host equipment.

Table 5-13: EUL Assessment by PA and Customer Agreement Date

Parameter Examined	PA EUL Documentation (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)							
	Pre-2013 Customer Agreement Date				2013+ Customer Agreement Date			
	PG&E	SCE	SCG	SDG&E	PG&E	SCE	SCG	SDG&E
Number of Measures Assessed	42	50	22	56	27	14	20	9
Percent of Measures with PA Tracking System EUL Populated	100%	100%	100%	100%	100%	100%	100%	100%
Percent of Measures with PA EUL Documented in the Project Application Files	33%	46%	41%	46%	4%	64%	40%	11%
Mean PA EUL Documentation Score	1.69	2.00	1.73	1.75	1.22	2.14	2.15	1.22
Median PA EUL Documentation Score	1.0	2.0	1.0	1.0	1.0	2.0	3.0	1.0
Summary of PA EUL Treatment								
Number of Measures with PA Tracking System EUL Populated (n)	42	50	22	56	27	14	20	9
Mean PA Tracking System EUL	13.55	13.12	13.82	14.04	12.96	12.79	10.45	13.89
Median PA Tracking System EUL	15.0	15.0	14.0	15.0	15.0	13.0	11.0	12.0
Number of Measures with PA EUL Documented in the Project Application Files	14	23	9	26	1	9	8	1
Mean PA Application File-Based EUL	12.93	13.50	12.11	13.81	20.00	14.44	11.25	15.00
Median PA Application File-Based EUL	15.0	15.0	12.0	15.0	20.0	15.0	8.5	15.0
Source of EUL								
DEER	7	11	5	13	--	4	3	--
Prior EAR Disposition	1	1	--	--	--	--	--	--
Workpapers	--	2	--	--	--	1	1	--
Prior PA Impact Evaluation(s)	--	--	--	1	--	--	--	--
Engineering Judgment	--	1	--	--	--	--	--	--
PA Study	--	--	--	2	--	--	--	--
Other	--	--	--	2	--	--	--	--
Source Not Specified	6	8	4	8	1	4	4	1

Table 5-13: EUL Assessment by PA and Customer Agreement Date (continued)

Parameter Examined	PA EUL Documentation (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)							
	Pre-2013 Customer Agreement Date				2013+ Customer Agreement Date			
	PG&E	SCE	SCG	SDG&E	PG&E	SCE	SCG	SDG&E
Summary of Evaluation EUL Treatment								
Number of Measures with Evaluation EUL Populated (n)	42	50	21	56	27	14	20	9
Mean Evaluation EUL	11.44	11.60	11.37	9.98	11.96	11.86	10.19	11.49
Median Evaluation EUL	12.5	12.0	10.5	8.0	15.0	11.0	11.0	12.0
Source of EUL								
DEER	19	29	9	36	16	9	13	2
Accept PA's EUL	2	3	3	1	--	--	1	--
Engineering Judgment	4	1	1	4	1	1	1	5
PA Study	--	2	--	--	--	--	--	--
Limited by the RUL of Host Equipment	4	4	3	4	6	1	1	2
Maximum Allowable EUL per CPUC (20 years)	--	--	1	1	--	--	1	--
Prior CPUC Impact Evaluation(s)	--	5	--	1	--	1	--	--
Prior EAR Disposition(s)	6	4	2	--	4	--	--	--
RCx/MBCx Program EUL Limit	6	2	--	5	--	--	--	--
Similar DEER Measures	--	1	--	1	--	--	--	--
Workpapers	--	--	--	--	--	1	1	--
Other	2	--	2	4	1	1	2	--
Summary of EUL Differences								
Number of Measures with Evaluation EUL Different Than PA Tracking EUL (n)	17	24	7	37	11	6	6	3
Mean Evaluation EUL (where differences exist)	7.39	8.79	6.43	7.34	8.63	11.33	11.12	9.47
Median Evaluation EUL (where differences exist)	6.0	7.0	6.67	8.0	8.0	12.5	10.5	6.7
Mean PA Tracking System EUL (where differences exist)	12.59	11.96	14.14	13.49	11.09	13.50	12.00	16.67
Median PA Tracking System EUL (where differences exist)	15.0	14.0	15.0	15.0	11.0	14.0	12.5	15.0

5.3.7 Calculation Considerations Assessment

Table 5-14 displays the calculation considerations used by the PAs and the ex-post evaluators in each time period. While the PAs relied on their own calculation tools more often than the evaluators, the PAs all do a good job of addressing a broad array of calculation considerations, in a manner that is generally equivalent to the evaluation process. However, the results do show that the PAs should strive to include CPUC policy and guidance, previous relevant EAR guidance, and standard evaluation practices more frequently than is their current practice.

Table 5-14: Calculations Considerations by PA and Customer Agreement Date

Parameter Examined	PG&E	Ex-Post M&V	SCE	Ex-Post M&V	SCG	Ex-Post M&V	SDG&E	Ex-Post M&V
Pre-2013 Customer Agreement Date								
Number of measures evaluated (n)	42		50		22		56	
Number of measures with calculation considerations documented (n)	40	40	49	49	20	19	45	54
Frequency of calculation considerations documented (n)								
CPUC policy and guidance	8	14	2	9	3	3	2	3
Previous relevant EAR guidance	7	9	2	10	2	7	2	13
Standard evaluation practices	8	11	4	18	3	6	8	22
Model calibration	11	8	7	9	--	--	7	5
PA calculation tools	15	10	12	5	5	4	6	3
Production normalization	7	6	9	12	3	3	1	1
Stable period of measured performance	6	13	12	15	5	4	5	3
Weather normalization	11	14	8	10	--	--	10	28
Seasonality considerations	5	6	5	5	--	1	3	2
Interactive effects	9	8	4	7	1	1	2	2
Facility-based custom elements/inputs	8	12	26	27	13	11	5	15
DEER inputs and assumptions	2	3	2	3	--	--	3	1
Pre-installation M&V	17	16	27	18	7	5	21	14
Post-installation M&V	11	17	27	26	5	6	31	26
2013+ Customer Agreement Date								
Number of measures evaluated (n)	27		14		20		9	
Number of measures with calculation considerations documented (n)	27	26	12	14	19	20	9	9
Frequency of calculation considerations documented (n)								
CPUC policy and guidance	--	3	--	3	--	1	--	--
Previous relevant EAR guidance	--	7	1	2	1	2	1	1
Standard evaluation practices	3	7	3	9	2	3	6	6
Model calibration	1	2	2	3	3	4	--	--
PA calculation tools	17	11	5	5	11	10	--	--
Production normalization	2	4	4	4	1	2	--	--
Stable period of measured performance	2	4	2	1	1	2	--	--
Weather normalization	4	4	--	--	1	1	--	--
Seasonality considerations	4	7	2	2	1	1	--	2
Interactive effects	5	3	3	1	--	--	--	--
Facility-based custom elements/inputs	15	15	3	3	10	12	--	1
DEER inputs and assumptions	--	--	--	--	--	--	--	--
Pre-installation M&V	14	13	7	4	4	3	9	8
Post-installation M&V	14	14	8	7	3	3	8	9

5.3.8 Calculations Assessment

Table 5-15, below, provides an assessment of the appropriateness, the documentation quality, and the accuracy of the PA models in determining measure savings. Model appropriateness scores in both the pre and post-2013 periods are between 1.98 and 2.78. These scores indicate that the calculation models used by the PA were appropriate but did not always properly consider key factors that may impact the savings such as weather/seasonality/production normalization (although these were generally taken into account by the PAs) and relevant CPUC or EAR guidance. PA model documentation quality received similar scoring (2.0 to 2.75), which suggests that some aspects of the PA models were not documented in enough detail to justify both pre- and post-installation conditions. Likewise, model accuracy ratings of 2.07 to 3.11 show that the PA calculation models were often deficient in using site-specific values for key parameters/variables or reliable and typical input variables such as flow rates, pressures, temperatures, and weather or production data.

Finally, Table 5-15 shows that in the pre-2013 period, the evaluator only used the PA model (or similar model) for 29 percent to 52 percent of measures. In all other cases, the evaluator used an entirely different model or deemed it necessary to make adjustments to the PA models. These figures improved considerably in the 2013+ period for SCE, SCG and SDG&E, where the evaluator used the PA model (or similar) for 61 percent to 100 percent of measures.

Table 5-15: Calculations Assessment by PA and Customer Agreement Date

Parameter Examined	PA Calculation Method Ratings (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)							
	Pre-2013 Customer Agreement Date				2013+ Customer Agreement Date			
	PG&E	SCE	SCG	SDG&E	PG&E	SCE	SCG	SDG&E
Number of Measures Assessed	42	50	22	56	27	14	20	9
Assessment of Appropriateness of PA Savings Model								
Number of Measures with PA Model Appropriateness Ratings (n)	41	50	22	55	26	14	20	9
Mean PA Model Appropriateness Rating	2.59	2.56	2.59	1.98	2.58	2.21	2.58	2.78
Median PA Model Appropriateness Rating	3.0	3.0	3.0	2.0	3.0	2.5	3.0	3.0
Assessment of Quality of PA Model Documentation								
Number of Measures with PA Model Documentation Quality Ratings (n)	41	50	22	55	26	14	20	9
Mean PA Model Documentation Quality Rating	2.54	2.64	2.55	2.00	2.46	2.07	2.75	2.56
Median PA Model Documentation Quality Rating	3.0	3.0	3.0	2.0	3.0	2.0	3.0	3.0
Assessment of Accuracy of PA Model								
Number of Measures with PA Model Accuracy Ratings (n)	41	50	22	55	26	13	20	9
Mean PA Model Accuracy Rating	2.54	2.48	2.59	2.07	2.58	2.38	2.65	3.11
Median PA Model Accuracy Rating	3.0	3.0	3.0	2.0	3.0	2.0	3.0	4.0
Assessment of Evaluation Use of PA Models								
Number of Measures with an Assessment of Evaluation Use of PA Models (n)	39	48	21	52	26	14	18	9
Evaluation used a different model	44%	25%	14%	46%	31%	21%	17%	--
Evaluation used a similar model	44%	52%	43%	29%	35%	71%	61%	100%
Evaluation adjusted the PA model	13%	23%	43%	25%	35%	7%	22%	--

5.3.9 PA Inputs and Assumptions Assessment

Table 5-21 summarizes the comprehensiveness, documentation, and accuracy ratings for the PAs' calculation method inputs and assumptions and provides an assessment of the evaluation team's use of the PAs' inputs and assumptions. As shown in the table, the mean comprehensiveness, documentation, and accuracy ratings for each PA are between 1.89 and 2.78, indicating that, on average, the PAs fell short of minimum expectations. No clear trends were seen between the pre- and post-2013 periods.

- PG&E's average ratings for the three inputs and assumptions evaluation categories ranged from 2.46 to 2.65 across the customer agreement periods; there were slight improvements in the comprehensiveness and accuracy ratings and a small decline in the documentation ratings from the pre-2013 to the 2013+ period.
- SCE's average ratings for the three inputs and assumptions evaluation categories ranged from 2.36 to 2.45 across the customer agreement periods; there were slight declines in the comprehensiveness and documentation ratings and an improvement in the accuracy ratings from the pre-2013 to the 2013+ period.
- SCG's average ratings for the three inputs and assumptions evaluation categories ranged from 2.14 to 2.70 across the customer agreement periods with notable improvements in all three ratings categories from the pre-2013 to the 2013+ period.
- SDG&E's average ratings for the three inputs and assumptions evaluation categories ranged from 1.89 to 2.78 across the customer agreement periods with notable improvements in all three ratings categories from the pre-2013 to the 2013+ period.

The evaluation team documented the use of different, similar, or adjusted inputs and assumptions compared to those used in the PAs' calculation methods. In the pre-2013 period, the evaluation team deemed it necessary to use different inputs / assumptions or to adjust PA inputs / assumptions for more than 50 percent of measures for all PAs. In the 2013+ period the evaluation team used similar inputs and assumptions as the PA more frequently than in the pre-2013 period, although changes / adjustments were still required for a 30 percent to 70 percent of measures.

Table 5-16: PA Inputs and Assumptions by PA and Customer Agreement Date

Parameter Examined	PA Calculation Input and Assumption Ratings (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)							
	Pre-2013 Customer Agreement Date				2013+ Customer Agreement Date			
	PG&E	SCE	SCG	SDG&E	PG&E	SCE	SCG	SDG&E
Number of Measures Assessed	42	50	22	56	27	14	20	9
Assessment of Comprehensiveness of PA Inputs and Assumptions								
Number of Measures with Ratings for Comprehensiveness of PA Inputs and Assumptions (n)	40	50	22	55	26	14	20	9
Mean PA Input and Assumption Rating	2.53	2.54	2.23	2.15	2.65	2.36	2.70	2.56
Median PA Input and Assumption Rating	3.0	3.0	2.0	2.0	3.0	3.0	3.0	3.0
Assessment of Documentation of PA Inputs and Assumptions								
Number of Measures with PA Input and Assumption Documentation Ratings (n)	41	50	22	55	26	14	20	9
Mean PA Input and Assumption Documentation Rating	2.59	2.50	2.23	2.24	2.46	2.36	2.65	2.78
Median PA Input and Assumption Documentation Rating	3.0	3.0	2.0	2.0	3.0	2.5	3.0	3.0
Assessment of Accuracy of PA Inputs and Assumptions								
Number of Measures with PA Input and Assumption Accuracy Ratings (n)	41	50	22	55	26	14	20	9
Mean PA Input and Assumption Accuracy Rating	2.49	2.36	2.14	1.89	2.50	2.43	2.65	2.67
Median PA Input and Assumption Accuracy Rating	3.0	3.0	2.0	2.0	3.0	3.0	3.0	3.0
Assessment of Evaluation Use of PA Inputs and Assumptions								
Number of Measures with an Assessment of Evaluation Use of PA Inputs (n)	40	45	21	50	23	14	20	9
Evaluation used different inputs	30%	29%	43%	42%	35%	29%	20%	--
Evaluation used similar inputs	40%	31%	24%	20%	17%	36%	55%	100%
Evaluation adjusted PA inputs	30%	40%	33%	38%	48%	36%	25%	--
Number of Measures with an Assessment of Evaluation Use of PA Assumptions (n)	35	42	17	50	25	13	20	9
Evaluation used different assumptions	34%	26%	41%	42%	24%	23%	20%	33%
Evaluation used similar assumptions	40%	48%	24%	38%	56%	31%	65%	67%
Evaluation adjusted PA assumptions	26%	26%	35%	20%	20%	46%	15%	--

6

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:

- 6.1 Gross Impact-Related Findings and Recommendations
- 6.2 Net-to-Gross-Related Findings and Recommendations
- 6.3 Program Findings and Recommendations Based on PPA Results

As summarized in Chapter 3, ex-post MMBtu lifecycle gross impact realization rates (LC GRRs) range by PA from 0.44 to 0.63.[55] Relative to 2010-12 custom impact evaluation results SCE, SDG&E, and SCG LC GRRs *decreased* by about 28 percent, 11 percent, and 15 percent, respectively on a combined MMBtu basis. LC GRR results for PG&E were 0.63 in both cycles. Net-to-gross ratios, ranging from 0.55 to 0.66, showed moderate improvement relative to 2010-12 custom impact evaluation results. Improvements in NTG were most pronounced for SCG, at 35 percent, while PG&E showed the least improvement (eight percent). At a summary level, the detailed recommendations in this chapter fall into the following primary areas:

- To more accurately estimate ex-ante savings, the PAs should:
  - Improve documentation and reporting of project EUL,[56] including a review of evaluation EUL conclusions/rationale in an effort to improve EUL claims and LC GRR results,

[55] Ex-post gross impact results were also developed in this evaluation for first year realization rates (FY GRRs), which range by PA from 0.54 to 0.75. Relative to 2010-12 custom impact evaluation results PG&E and SDG&E FY GRRs *increased* by about 15 percent and 73 percent, respectively. FY GRR results *decreased* for SCE by about 10 percent and three percent for SCG. It is notable that increases in FY GRRs are an indication of improvements in ex-ante engineering-based savings estimates and associated PA process improvements, whereas increases in LC GRRs are an indication of improvement in a combination of engineering-based savings estimation and EUL and early retirement (ER) treatment (including associated RUL and EUL considerations). Relatively good PY2013 FY GRR results for each of the PAs were reduced when taking into consideration EUL and ER factors supporting LC GRR results. This was especially true for SDG&E, with a PY2013 FY GRR of 0.75 and an LC GRR of 0.49.

[56] It is notable that the evaluation estimate of EUL differed from the PAs estimate 46 percent of the time. For those instances the evaluation-derived average EUL was smaller than the ex-ante average EUL by more than 4 years, representing a 35 percent reduction in the ex-ante EUL claim for that subset of observations. As noted in Chapter

 - Improve quality control of determining project operating conditions, ex-ante baseline determinations, savings calculations, and eligibility rules to address the discrepancy factors presented in this report, and
 - Ensure adjustments to project savings based on post-installation inspections and M&V.

- To achieve sufficient quality control, PA project documentation needs to significantly improve. SDG&E appears to be leading the other PAs in this area, given the level of improvement noted in Chapter 5 in 2013+ applications.
- To reduce continued moderate free ridership, PAs should test changes to program features designed to increase program-induced savings.

6.1 Gross Impact-Related Findings and Recommendations

As presented in Chapter 3, it was found that operating conditions, inappropriate baselines, calculation methods, and ineligible measures were all important discrepancy factors which contributed to impact-related differences between ex-post evaluation results and utility savings claims. Program improvements in these four 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:

- Projects Should Save Energy
- Operating Conditions
- Baseline Specification
- Calculation Methods

6.1.1 Projects Should Save Energy

All PAs had projects with negative and/or zero GRRs, and these served to lower the weighted realization rate considerably. Out of 189 M&V points, 31 projects, or 16 percent of the sample, had a GRR of zero or lower. The discrepancy factors that led to these low realization rates were identified in Chapter 3, and 23 of the cases were due to one of two factors – inappropriate baseline or ineligible measures.

---

3 LC GRR results were substantially lower than FY GRR results and this EUL difference was *the key factor* driving down the LC GRR results. As described in Chapter 5, primary drivers of EUL differences include different interpretations of applications of DEER, EUL limited by RUL of host equipment for add-on / systems optimizations, and RCx program rules.

There is clearly a need for the PAs to improve in the areas of estimation accuracy and quality control for all projects, but in particular there is a need to focus on projects where the ex-post savings are zero or even negative. Baseline selection and eligibility screening are pretty basic steps in the development of ex-ante savings estimates and represent relatively easy-to-implement areas for improvement. ***Recommendations include the following:***

- ***Improve PA program requirements, manuals, training, and quality control procedures in order to screen out ineligible projects***. A more thorough PA 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.
- Regarding eligibility, the evaluation team recommends that the PAs clearly document the energy efficiency action that is being performed and ensure the installed measures meet program payback requirements. As recommended in the previous evaluation cycles, the PAs should adjust the set of qualifying measures/technologies that are eligible for incentives and annually review the list of qualifying measures for each program to eliminate eligibility for those that became standard practice.
- Furthermore, ***it is recommended that the PAs carefully review each of the 31 FSRs listed in Table 3-6 to identify the specific reasons that led to zero or negative savings, and use those lessons learned to improve related project practices.*** An array of different factors led to very low site-level GRRs, but some common reasons include: like-for-like replacement of equipment, improper application of ISP, baseline changed from ER to ROB/NR, improper application or interpretation of code requirements, and failure to apply the non-regressive baseline rule.
- It is recommended that the PAs make greater efforts to address the same types of projects that received low GRRs in this evaluation, given the significant downward effect that these projects had on the ex-post gross savings estimate.

6.1.2 Operating Conditions

The operating conditions discrepancy factor accounted for 29 percent of all downward evaluation GRR result adjustments. Some aspects of operating conditions estimation can be addressed through improvement in program implementation activities and quality control.

<u>Finding: Changed Operating Conditions for Projects</u>

Evaluated operating conditions were often found to be different than described in program project documentation. Per evaluation guidelines, measures are evaluated as-found, and the ex-post savings analyses were performed for the as-observed/verified conditions, including back-casting where relevant to current operations and did not include any forecasting.

- **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 PAs should ensure the use of site-specific inputs whenever possible. Also assumptions used should reflect conservative values supported by strong evidence from secondary sources.

  PAs 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. PA 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 parameters.

- The evaluation team recommends that the PAs ensure that savings calculations are based on actual equipment-use schedules and reflect any changes to the post-installation operating parameters (such as flow rates, temperatures and set points, system pressures, production rates, and power measurements). The PAs should always include a quality control check on equipment operating hours, especially identifying any idle periods and removing those as necessary in the ex-ante energy savings models.

  Consideration should be given to selecting an appropriate and representative time period to use for data collection and savings determination. Increased use of selective parameter measurement using uncertainty analysis and short term monitoring is also recommended.

- Another key issue is that evaluators discover that the production period observed in the post-retrofit case is often too short (one week or less) and not typical of the production or operating variations that the equipment will be subject to over the course of a year. To help mitigate this issue, the PAs should wait for measure operation to stabilize and become typical prior to truing-up the ex-ante models and making a savings claim.
- As stated in previous evaluation cycles, the PAs should use longer-term pre- and post-installation M&V activities and true-up the savings estimates to reflect most recent measure operation. The PAs should also normalize for production fluctuations between pre- and post-installation periods.

  In some cases, PAs should delay claiming energy savings for projects if the installation is not complete or if operations are very unstable or unrepresentative of expected ex-post conditions. The PAs 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.

- For projects entailing the use of simulation models, the evaluation team recommends that these models be re-run after the equipment is commissioned and building loads represent steady state operation.
- For pump efficiency improvement projects, use actual pump efficiency tests, not estimates.

  PAs 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 revisions. PAs should delineate expectations for post-retrofit inspection paperwork and require inspectors to identify, collect and record pertinent measure operating parameters, as well as quantities in both pre-installation and post-installation efforts. PAs should consider holding multiple trainings, regularly (e.g., quarterly), with internal staff, implementers, and PA 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 thereby work to ensure comprehensive and consistent M&V practices well beyond a cursory verification that new equipment was present at a given site.

6.1.3 Baseline Specification

Improper baseline specification resulted in a substantial number of adjustments, resulting in significant impacts to ex-ante savings claims for both electric and gas projects. These adjustments arose from improper project baseline specification, improper baseline operation, or crediting new or replacement equipment with improved efficiencies when, in reality, the new equipment efficiency did not exceed industry standard practice.

<u>Finding: PA Baseline Changed by Evaluation</u>

There was generally good agreement on project type and project baseline when comparing PA and evaluator selections. ER and ROB projects were the most commonly overturned project types across all PAs.

- **Recommendation**: ***Increase efforts to*** e***nsure conformance with CPUC baseline policies and make a greater effort to examine existing equipment RUL.*** The PAs should mount a concerted effort to adopt baseline specification practices in conformance with Decision 11-07-030 and CPUC policy. Conformance with these guidelines and accurate specification of project baseline type, such as early retirement, normal replacement, replace on burnout, system optimization, new construction, and add-on measure would eliminate many of these issues. The PAs should amend 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). If the efficiency of the pre-existing equipment is higher than the replacement equipment baseline, then the PAs should select the pre-existing equipment as the baseline.

- PA RUL documentation was found to be significantly lacking. For appropriate selection of baseline, RUL assessment is needed for all projects except capacity expansion and new construction projects. For example, RUL assessment of add-on projects is used to examine the expected remaining life of the host equipment, for the purposes of setting EUL for the add-on measure. Other examples of applicability of RUL are more straight-forward and therefore not worthy of further mention here. For all early replacement (ER) projects, the PAs should provide and clearly document the remaining useful life (RUL) of the pre-existing equipment. It is recommended that the PAs carefully review the evidence collected to estimate the RUL for all early retirement applications. The PAs must also conduct appropriate due diligence to insure that for an ER project the current removed system would be able to meet the service requirements of the newly installed program equipment and that failure of the replaced equipment is not imminent.
- **Recommendation**: ***Clearly identify project event in terms of natural replacement, replace on burnout, early replacement, new construction, and add-on equipment, 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 PAs 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 PAs' 2013 procedures manual.*
- **Recommendation**: ***Disseminate information on baseline selection to ensure best practices across program staff, implementers and customers.*** The evaluation team recommends that the PAs should provide their program staff, implementers and customers with the most current industry standard practice (ISP) studies and the CPUC's guidance documentation. This will help better align the PA's baseline selection with the CPUC's directives.

<u>Finding: Greater PA Effort Needed for Proper Baseline Selection</u>

Choosing a proper baseline requires systematic examination of a number of factors. Evaluation efforts led to a number of cases where PA baseline selection was overturned.

- **Recommendation**: ***The PAs need to do a better job of ensuring that baseline equipment specifications are capable of meeting post-installation operating requirements, that the baseline selected is consistent with the project type, and that regressive baseline***

considerations are examined. The evaluation team recommends that for all capacity expansion projects, the PAs ensure that the baseline equipment meet the post-install operating and production capacities. In-situ equipment (unless it is above code or ISP) is an invalid baseline to calculate energy savings for normal replacement (NR), replace-on-burnout (ROB), capacity expansion and new construction (NC) projects. Additionally, the evaluation team recommends that the PAs carefully review projects for possible regressive baselines and document the pertinent findings.

6.1.4 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 6.1.2. Recommendations to improve calculation methods and protocols are provided below.

Finding: Inadequate Impact Methods and Models

Inadequate or suboptimal methods, models, and inputs were observed in the M&V sample. The evaluation used a different model than the PA in roughly 30 percent of projects included in the evaluation gross impact sample. The evaluators often found it necessary to modify PA models, inputs and assumptions. In some cases, the PA did not properly take into account key factors that may impact the savings such as weather/seasonality/production normalization. Generally models needed to be adjusted because the PAs did not properly account for CPUC policy and guidance, previous EAR guidance, and standard evaluation practices.

- **Recommendation:** ***Continue to review and improve impact methods and models through review of evaluation results, industry best practices, and collaboration with the CPUC's ex-ante review process.*** The PAs 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. PAs 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 PAs 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.
- The evaluation team recommends that the PAs provide their implementers and/or customers with the most current, standardized or CPUC-approved calculation tools.

- Further, the evaluation team recommends that the PAs include in each application file the live, unlocked, non-password protected spreadsheet models. For projects entailing simulation models, the PAs should record key model inputs and outputs, in addition to providing the final analysis spreadsheets/models.
- **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 PA technical staff prior to finalization of incentives and savings claims. These reviews by knowledgeable technical staff can help ensure reliable and accurate impact estimation.
- **Recommendation:** ***Conduct periodic due diligence to ensure programs adhere to PA 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 PAs 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 PAs 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 third party programs.

<u>Finding: PA Models Were Not Always Calibrated Using Observed Conditions</u>

Key inputs and observations, when available, based on ex-ante field verification, installation reports and M&V, were sometimes not subsequently incorporated within the ex-ante impact models.

- **Recommendation:** ***The PAs should true-up savings based upon post-installation data, such as by calibrating the simulation model to utility usage data.*** For example, for large energy end-uses, the evaluation team recommends that the PAs use pre-and post-installation billing or AMI data (if available) as a sanity check or to better calibrate the actual energy savings. The PAs should also make better use of available post-installation M&V power measurements, whether spot readings, short-term or long-term interval data.
- **Recommendation:** ***For projects entailing weather sensitive measures, the PAs should verify that the correct weather files are used in the analysis. Regarding peak demand analysis, adopt CPUC protocols and procedures as they relate to the California climate zone peak period definition.*** Peak impact estimates should reflect loads during the California climate zone three-day period, if data was collected during the actual three-day peak for that region or during the peak summer time period of 2-5pm from June 1 through September 30. Calibration considerations noted above apply also to peak, including the use of post-installation M&V power data that best represents the coincident peak period.

6.2 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 Chapter 4 of this report.

Finding: Free Ridership for Custom Projects Remains Elevated

On a statewide basis, the NTGR averaged 0.54. NTGR results indicate a medium[57] level of free ridership and a resulting medium level of program influence. Although this demonstrates some improvement since PY2010-2012, this value continues to be similar in magnitude to NTGRs from the past several evaluation cycles, as shown in Table 6-1. The general conclusions are that free ridership has not changed substantially for custom programs. 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 appear to have been taken by the programs with the specific goal of reducing free ridership.

Table 6-1: Statewide Industrial Custom Program[58] Evaluation Net to Gross Ratios, Program Years 1998-2013

(1 – Free Ridership)	1998	1999	2000	2001	2002	2004-2005	PY2006-2008		2010-2012	2013
							PG&E	SCE		
Weighted	0.53*	0.51	0.41	0.65	0.45	0.57	Electric - 0.45, Gas - 0.31	0.63	Electric - 0.48, Gas - 0.53 MMBtu – 0.50	Statewide MMBtu - 0.54

* Weighted by incentives rather than by kWh savings.

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 primarily 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, program influence was very low. There were also

57 Medium free ridership is defined in this report as between 25 percent and 50 percent (i.e., NTGR of between 0.50 and 0.75).

58 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.

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 PAs 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.
- **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 have become standard practice. At a minimum, such reviews should take place annually. Measures that are already likely or very likely to be typically installed should 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.

  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 new construction approach and requiring the project to reach a minimum savings threshold (such as 15 percent) 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 product 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. Although it is certainly true that many customers do not adopt attractive efficiency projects with very low paybacks,[59] 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 comprehensive mix of program features and leverage an array of delivery channels to engage individual customers and encourage a long-term energy efficiency-based focus.*** Use a broad mix of program features and delivery channels to market projects and encourage deeper impacts over time. In addition to incentives, make appropriate use of 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 an appropriate mix of customer outreach and delivery choices.
- **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. There is inadequate financial analysis conducted to determine what portion of the customer's financial investment threshold is associated with the energy

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

savings of particular projects versus non-energy factors[60] such as increases in production and reductions in labor, 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.

6.3 Program Findings and Recommendations Based on PPA Results

Findings and recommendations stemming from the project practices assessment activities are presented in this section. PPA findings and recommendations are presented in the following subsections:

- Overarching Considerations
- Project Eligibility
- Project Type and Related Baseline Selection
- Project Cost-Effectiveness Treatment

6.3.1 Overarching Considerations

The PPA sample design included two time period segments – applications with a customer agreement date falling before 2013 (pre-2013) and for 2013+. The purpose behind this design was to isolate and report separately on the 2013+ time period. In this design the 2013+ period represents custom project applications under full EAR influence. This is asserted, given that the EAR process began in earnest in January 2012.

Finding: Inadequate Impact Methods and Models

For three of the PAs the PPA assessment found limited evidence of improvement in PA performance in the 2013+ period relative to pre-2013. SDG&E was an exception. Although the sample size was limited for SDG&E for the 2013+ period, consisting of just 9 points, SDG&E performance improved in all of the following areas:

[60] 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.

- Showed improved eligibility ratings
- All 9 projects were determined to be eligible
- The project type and related baseline definition were both correct 100 percent of the time
- Project cost data scores showed great improvement and SDG&E is the only IOU documenting and recording incremental project costs in the 2013+ time period
- Led scores on incentive estimation
- Best on model accuracy scores, and the evaluation used a similar model for all 9 projects
- Demonstrated improved handling of inputs and assumptions

For the other PAs the appearance is that this is still a new process, and PAs may have not yet disseminated relevant guidance throughout their organizations. The 2013 PPA results, combined with GRR and NTGR findings, provide a solid baseline from which to continue tracking PA performance.

- **Recommendation:** ***It is recommended that a statewide document, similar to the PPA form, be developed for use by all PAs for custom claims.*** The PPA forms developed by the evaluation team provide a very structured and methodical way of examining energy efficiency measure claims. The PAs go through a similar process but in a less systematic way, and improvements to forms and processes should have a positive outcome on results. The evaluation team believes that this approach will help PAs improve their GRRs and documentation, especially through more careful consideration of first-order factors affecting project eligibility and project baselines.

6.3.2 Project Eligibility

Overall, less than half of measures in both the pre-2013 and 2013+ periods had eligibility considerations documented by the PAs.

Finding: The Evaluation Identified Projects with Eligibility Issues Including those Related to Baseline Selection

The large majority of measures in the 2013 evaluation were "routine" and thus required minimal documentation to support project eligibility.[61] However, there were a number of cases where ISP or code-based baseline determination rendered a project ineligible. In these cases where project

[61] See discussion of routine and non-routine measures in Chapter 5, section 5.3.1.

eligibility and baseline are directly linked, documentation must be sufficient to establish above code/ISP performance, even for "routine measures."

- **Recommendation:** ***The PA's project eligibility treatment suggests that the PA's communication and coordination efforts with entities responsible for implementing CPUC guidance should be increased.***
- **Recommendation:** ***The requirement that measures exceed the ISP / code baseline is a first order consideration for project eligibility***. As such, it is important that the PAs spend adequate time documenting the appropriate project type and project baseline when establishing eligibility.

6.3.3 Project Type and Related Baseline Selection

Project type and baseline analyses showed relatively good agreement between PA and evaluator designations.

Finding: ER and ROB projects were the most commonly overturned project types across all PAs, and PA existing equipment baselines were frequently over-turned by the ex-post evaluation.

Inadequate PA documentation quality surrounding the age and condition of existing equipment, EULs/RULs, capabilities of existing / baseline equipment to meet service capabilities, and available efficiency levels likely contributed to different conclusions by the PAs and evaluators regarding ER, NR, and ROB project types. Similarly, the evaluation team more often examined the efficiency level of replaced equipment which helped to identify instances of regressive baselines and subsequent project ineligibility. Finally, for some PAs, evaluators more often reviewed ISP and applicable codes and standards, which were some of the more common reasons for overturned project baselines.

- For applicable recommendations, refer section 6.1.3 above.

6.3.4 Project Cost-Effectiveness Treatment

Project cost-effectiveness variables and those associated with estimation of lifecycle savings receive inadequate IOU attention in custom application files.

<u>Finding: EUL documentation is maintained in less than 50 percent of custom application files. Where directly relevant, such as for ER claims, RUL is generally not documented in custom application files. Where directly relevant, incremental cost is only documented in custom application files about 40 percent of the time.</u>

The EUL should always be recorded in the in the post-installation inspection report. For ER measure RULs, neither the tracking system nor project application files were typically populated. RUL values for ER measures should always be populated in order to accurately assess savings in dual baseline situations.

Incremental cost values for applicable measures were especially deficient and were populated only for 45 of 104 measures (43 percent). For cost effectiveness purposes, incremental cost values must be populated and clearly documented for all ER, ROB, NR, NC, and capacity expansion projects.

- **Recommendation:** Better ex-ante documentation is needed supporting these and other important project parameters. The trend for needed improvement in this area is true for each of the IOUs, except for the most recent SDG&E trends noted above. For the 2013+ period SDG&E documentation supporting key project conclusions, parameters, inputs and assumptions improved significantly.