Itron
Knowledge to Shape Your Future

Electric | Gas | Water
information collection, analysis and application

Final Report

2014 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
DNV GL

CALMAC Study ID CPU0147.01

April 29, 2016

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) 2014 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' 2014 efficiency programs and was conducted under the Industrial, Agricultural and Large Commercial (IALC) Roadmap as part of an overarching contract for PY2013-2015 evaluation services.

The evaluation addresses custom, non-deemed measure installations, and the scope includes a variety of projects that received incentives via more than 100 utility programs.[3] The scope of work for the evaluation of custom measures includes an independent estimation of gross impacts (i.e., evaluated savings realized from the project) and net impacts (i.e., evaluated gross savings adjusted to account for savings attributable to the program),[4] and a Project Practices Assessments (PPA) activity[5] to discern possible changes in ex-ante savings development practices. Findings and recommendations to improve program performance are also provided.

California PA-led custom programs are currently evaluated on an annual basis. The final version of the data necessary for the 2014 IALC evaluation was received by October 29th, 2015. The majority of the evaluation field work and data collection activities took place in the second half of 2015, and the analysis and reporting was completed in the first half of 2016. Evaluation of the 2015 programs is beginning in the first half of 2016. This annual evaluation schedule results in

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 the CPUC under the direction of CPUC staff responsible for evaluation of utility energy efficiency programs.

3 Custom projects are those where the energy savings are calculated specifically for the individual project; deemed measures have designated savings that apply to various categories of projects and are not calculated specifically for each site.

4 The reader is referred to Appendix G for a glossary of common terms used in this evaluation report.

5 Project Practices Assessment reviews were conducted for all completed measurement and verification (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.

faster feedback to the PAs with regard to their program activities and supports the Efficiency Savings and Performance Incentive (ESPI) award.[6]

Three main evaluation activities support the findings and recommendations in this report: (1) M&V activities for estimating gross impacts for 150 projects, (2) telephone survey data collection supporting net to gross (NTG) estimation for a total of 196 projects, and (3) a total of 150 engineering reviews supporting PPA results.

1.1 Custom Impact Evaluation Portfolio Context and Sample Sizes

The programs included in this custom impact evaluation carry out energy efficiency projects at a wide range of nonresidential locations, including commercial, institutional, agricultural and industrial facilities. The scope of this evaluation addresses nonresidential custom measures of all types with two main exceptions: custom lighting measures and pump test claims.[7] Additionally, 103 heating, ventilation, and air conditioning (HVAC) records were assigned to the IALC roadmap after 2014 field work was completed and were therefore not evaluated.[8] Each custom-oriented PA program offers one or more of the following interventions to encourage end users to upgrade to energy-efficient measures or improve processes: site-specific facility assessments/audits, feasibility studies, project incentives, pump testing, and specialized training.

In 2013, the IALC custom evaluation was divided into two Work Orders in an effort to isolate and report separately on non-residential whole building new construction projects. For PY2014 these two impact evaluations were combined into one custom impact evaluation. Under the current custom evaluation, NRNC whole building projects are treated like any other points in the program population or evaluation sample. That is, NRNC whole-building project results are pooled with other projects to derive aggregate impact results, including gross realization rates (GRRs) and net-to-gross ratios (NTGRs).

As shown in Figure 1-1, energy savings claims associated with the scope of this evaluation represent a significant contribution to the overall savings claims for the PAs' energy efficiency programs, accounting for about 24 percent of statewide electric savings claims and 61 percent of

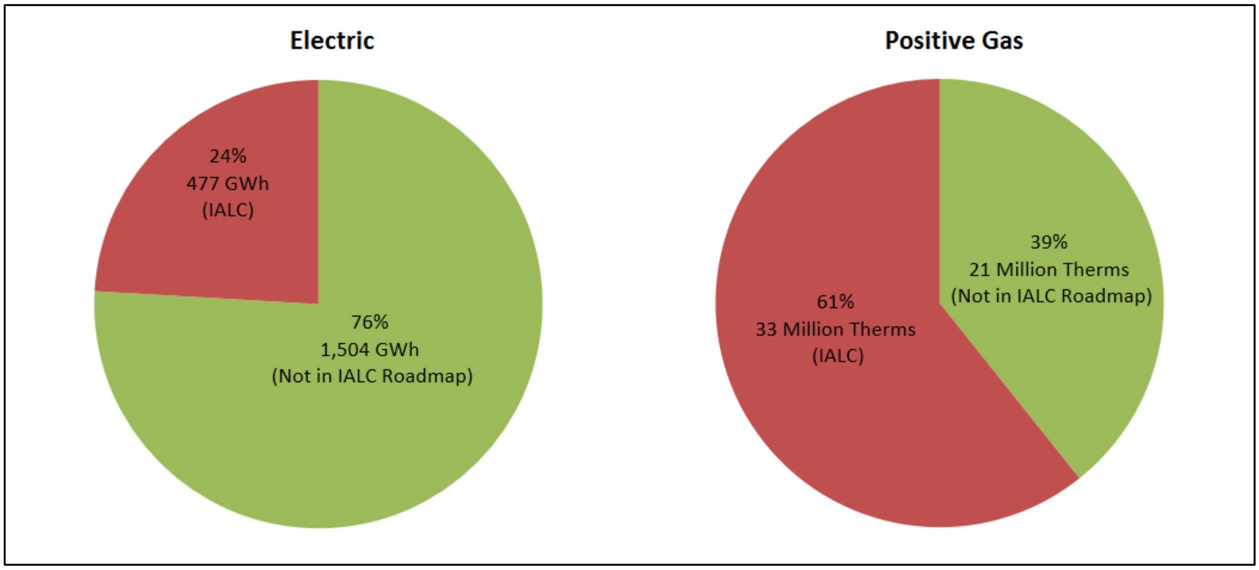
[6] CPUC Decision 13-09-023 established the ESPI mechanism, which awards PAs for performance in both resource and non-resource activities supporting energy efficiency.

[7] Custom lighting measures are addressed in a separate impact study on nonresidential lighting, under the CPUC 2013-2014 Commercial Roadmap. Pump test claims were evaluated in the past and were not in scope for 2014.

[8] Any savings from these 103 HVAC claims were passed through (i.e. PA savings were accepted as claimed). An additional 10,153 HVAC records from PG&E were also assigned to the IALC roadmap after 2014 fieldwork was completed. These records correspond either to HVAC diagnostic or maintenance measures, or to "contract signing" measures (program PGE21015). These claims receive incentives but claim zero savings and also were not evaluated.

statewide gas savings claims[9] during PY2014. 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 477 GWh and annual gas savings claims totaling 33 million Therms.[10]

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



* "Positive" gas refers to the 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 domain, the primary sampling domains for developing and reporting gross and net impact results were by each PA territory on a combined MMBtu basis,[11] where applicable for that PA. This approach resulted in the following four sampling domains for which gross realization rates[12] and

9 Excluding negative savings impacts associated with HVAC interactive effects.

10 Of the 477 GWh and 33 million Therms associated with the IALC Roadmap, 15 percent of electric claims and nine percent of gas claims were attributed to NRNC whole building projects. This equates to approximately 12 percent of IALC savings on a combined MMBtu basis.

11 MMBtu is a measurement of energy that means one million British Thermal Units (Btus) and is a way of expressing total energy from both the electric and gas savings. 1 MMBtu =1,000,000 Btu, 1 Therm = 100,000 Btu source energy, 1 kWh = 10,239 Btu source energy. Conversion rates obtained from "*2001 Energy Efficiency Standards for Residential and Non-residential Buildings, California Energy Commission*," June 2001.

12 "Gross realization rate" is the evaluation gross results/PA savings claims.

net to gross (NTG) ratios[13] 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 150 gross impact points (consisting of 192 individual measures) and 196 net impact points, where a point or "sampling unit" is defined as an individual project (from one or more records) installed at a specific site. The original sample design was to have 200 NTG project surveys completed, 150 of which were to overlap with the 150 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, 116 of the completed NTG sample points overlapped with the 150 evaluated gross M&V points. 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	43	0	42	10	31%	37%
SCE	39	3	46	11	30%	29%
SDG&E	34	1	34	11	49%	43%
SCG	29	1	35	7	55%	38%
All PAs	**145**	**5**	**157**	**39**	**36%**	**36%**

1.2 High-Level Custom Gross Impact Results

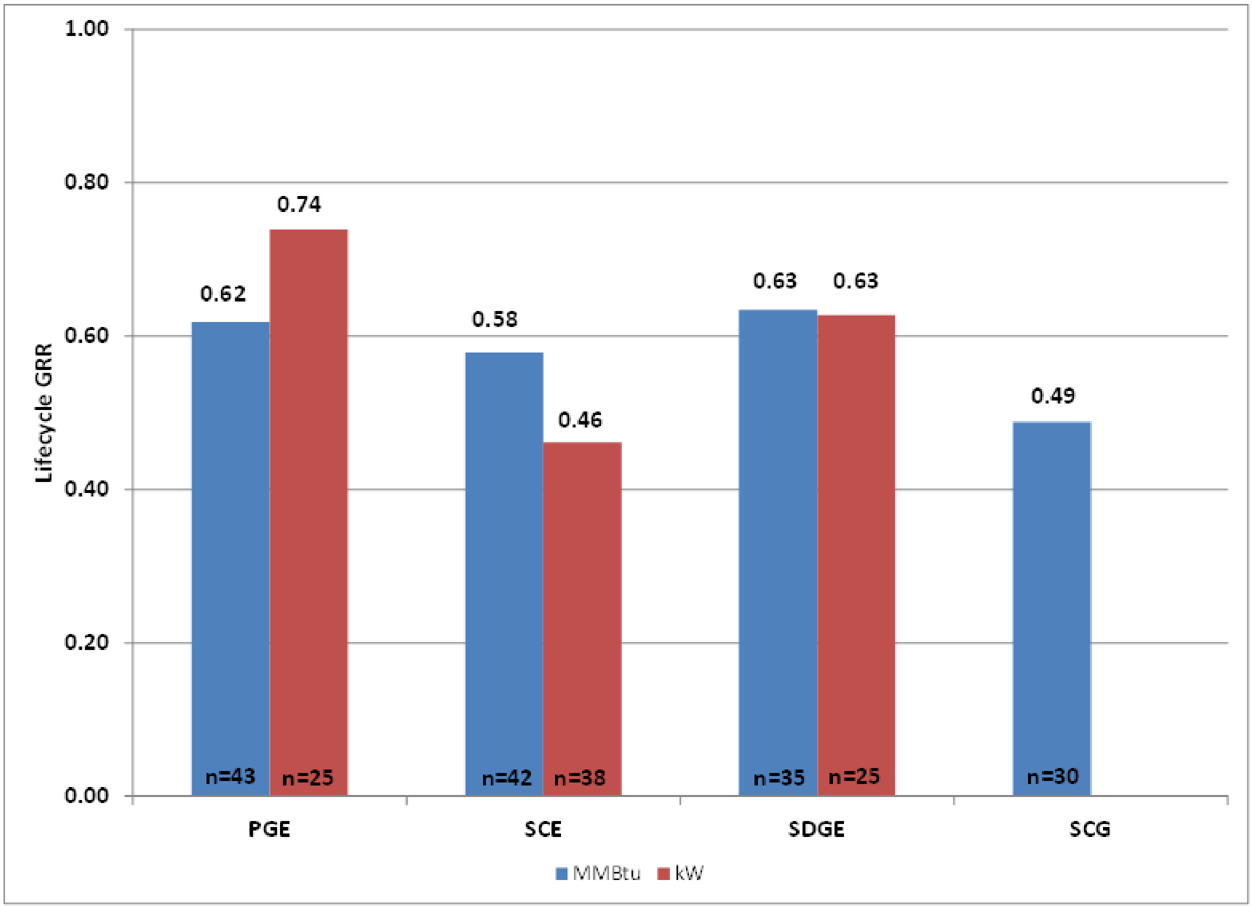
Figure 1-2 and Table 1-2 below summarize the mean lifecycle gross impact realization rates (GRRs) for each of the four PA sample domains. Gross realization rates are calculated for each sampled project as the ex-post, evaluation based estimate of savings divided by the PAs' ex-ante estimate of savings. 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 PG&E kW is 0.74) and are similar to those from the 2010-2012 and 2013 evaluations (see weighted MMBtu comparison in Chapter 3). For all PAs except PG&E,

[13] Net to Gross (NTG) ratios are used to estimate and describe the "free ridership" that may be occurring within energy efficiency programs, that is, the degree to which customers would have installed the program measure or equipment even without the financial incentive (e.g., rebate) provided by the program.

weighted lifecycle realization rates are lower than the corresponding first year realization rate.[14] Generally, evaluation lifecycle realization rates remain significantly below the 0.9 default ex-ante GRR adjustment for custom programs. A significant number of projects (22 out of 150) 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* and the *2013 IALC Custom Impact Evaluation Final Report.*[15] In the 2014 sample, the relatively high standard deviation 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,

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

[15] http://www.calmac.org/publications/2010-12_WO033_Custom_Impact_Eval_Report_Final.pdf
http://www.calmac.org/publications/IALC_2013_Report_Final_071715.pdf

which will always result in a large standard deviation regardless of the number of projects sampled. For example, the M&V sample supporting PG&E's kW GRR consisted of 25 projects, five of which had LC GRRs of greater than 1.5 and seven projects with LC GRRs of zero or less, resulting in an error ratio of 1.67. While the precision of the 2014 results are similar to previous evaluation results, the reader should be cognizant of the relatively broad confidence intervals when interpreting the results and findings.

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

Energy Metric	Population Count	Sample Count	% of Savings Sampled	LC Mean GRR	Error Ratio**	90% Confidence Interval	FY Mean GRR
PG&E ŧ							
MMBtu*	1,244	43	31%	0.62	0.75	0.50 to 0.73	0.59
kW	942	25	14%	0.74	1.67	0.34 to 1.14	0.69
SCE							
MMBtu*	1,161	42	30%	0.58	0.94	0.44 to 0.71	0.64
kW	1,069	38	25%	0.46	1.00	0.34 to 0.58	0.50
SDG&E							
MMBtu*	203	35	49%	0.63	0.43	0.57 to 0.70	0.73
kW	139	25	38%	0.63	0.45	0.54 to 0.71	0.67
SCG							
MMBtu	236	30	55%	0.49	0.96	0.36 to 0.62	0.58

* The primary sample was designed and selected based on ex-ante MMBtu savings estimates. 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.

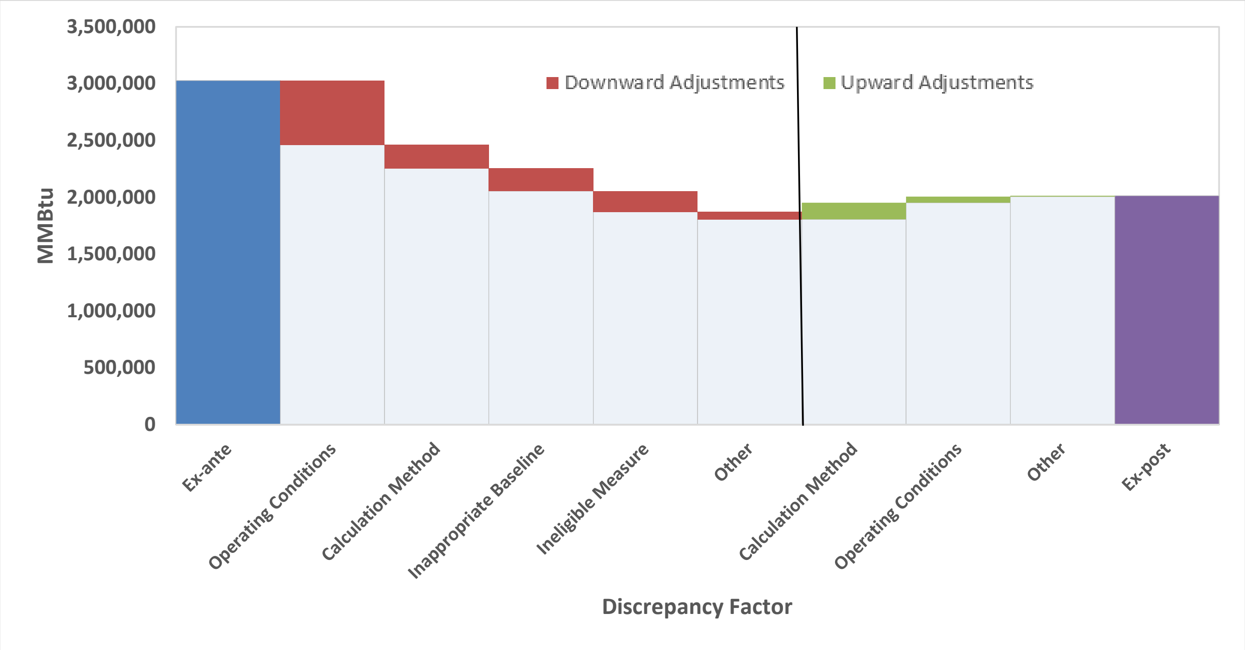
ŧ Unlike the other PAs, PG&E LC GRR results are higher than FY GRR results. This is caused by five PG&E projects with significant weights in their respective strata that were assigned a *higher* EUL by the evaluation. This caused the LC GRR of those strata to increase compared to the FY GRR, and ultimately caused a slight increase in the weighted LC GRR as compared to the weighted FY GRR.

The four principal reasons that ex-ante gross impacts differ from ex-post results are: (1) use of observed operating conditions, (2) the PAs' calculation methods, (3) baseline specification, and (4) ineligible measures. These discrepancy factors were examined for all projects where they caused upward or downward adjustments to the ex-ante savings.[16] Of the 192 records (measures) studied, these discrepancy factors explain a portion of the downward adjustments in ex-ante savings for 39

[16] Factors that led to downward adjustments are examined more thoroughly in this report in order to best inform improvements to ex-ante savings estimates, given that average gross impact realization rates are far below 1.

percent, 29 percent, 16 percent and 10 percent of records, respectively.[17] For all downward adjustments, across all PAs, these four factors reduced the sample-wide ex-ante MMBtu savings estimates by a combined amount of 38 percent; 19 percent (operating conditions), 7 percent (calculation methods), 7 percent (inappropriate baseline), and 6 percent (ineligible measures). Figure 1-3 depicts, by discrepancy factor, all downward and upward adjustments that were made to the ex-ante MMBtu savings estimates for the sample of 2014 M&V points.

Figure 1-3: Summary of Discrepancy Factors Resulting in Downward and Upward 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.

[17] More than one discrepancy factor often applies to a given record. Other reasons for differences in savings results were observed less frequently, but include the following: inoperable measures, incorrect measure counts and tracking database discrepancies, among others.

Figure 1-4: Weighted Net-to-Gross Ratios by PA[18]

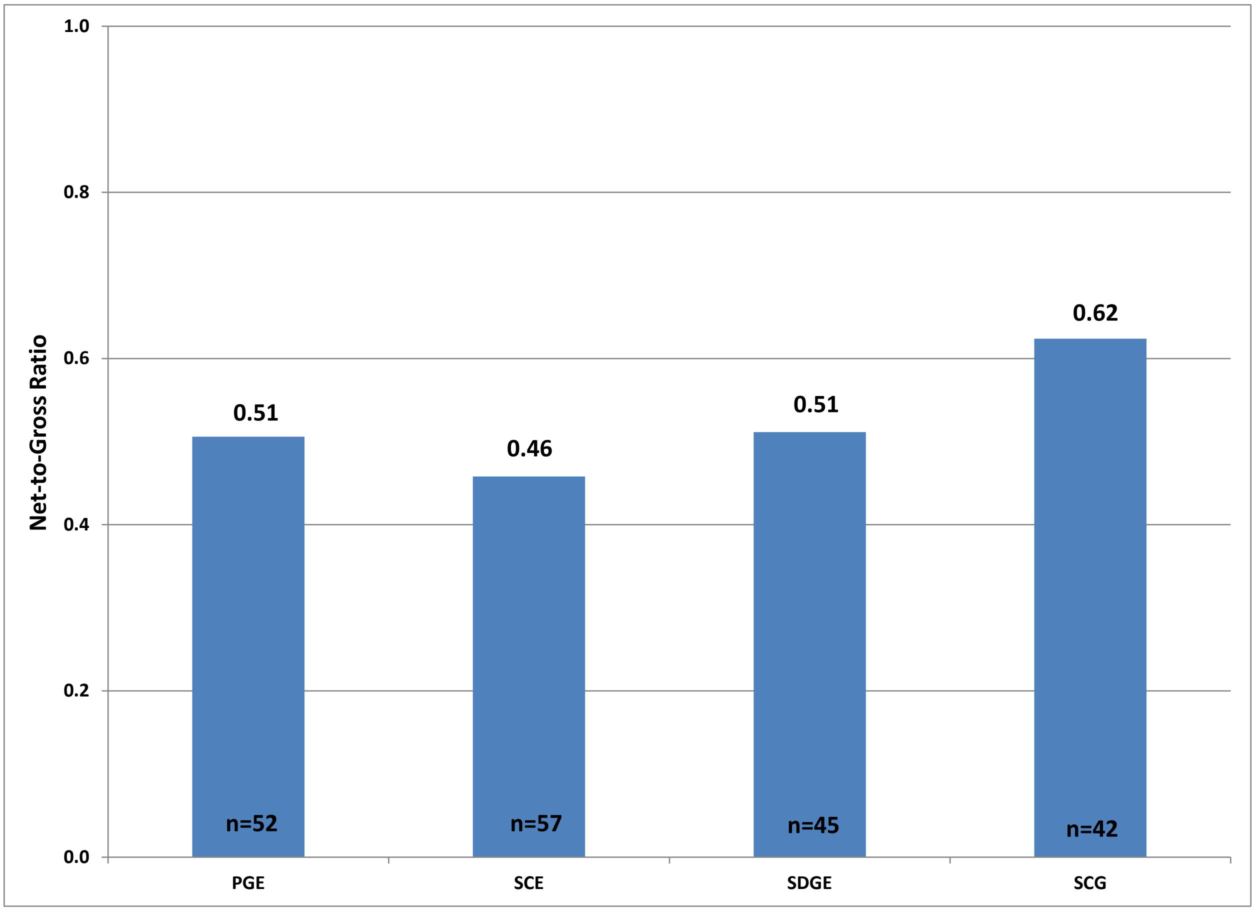


Table 1-3: Weighted Net-to-Gross Ratios by PA

Results	Mean Net-to-Gross Ratios			
	PGE	SCE	SDG&E	SCG
Weighted NTGR	**0.51**	**0.46**	**0.51**	**0.62**
90 Percent Confidence Interval	0.47 to 0.54	0.42 to 0.49	0.47 to 0.56	0.6 to 0.65
Relative Precision	0.06	0.07	0.09	0.04
n NTGR Completes	52	57	45	42
N Sampling Units	1,244	1,161	203	236
Error ratio (ER)	0.29	0.35	0.42	0.18
Percent of Ex-Ante MMBtu Savings	37%	29%	43%	38%

[18] Note that these values reflect the removal of 4 projects from NTG calculations. As described in Chapter 4, this was due to either an ineligible measure (3 projects removed) or inconsistent answers between the EAR and post-installation survey responses (1 project removed).

Based on the NTGR results presented above, the following observations are noteworthy:

- At the level of PA sampling domain, the final NTGRs range from 0.46 to 0.62, showing some decline compared to the previous cycle (see weighted MMBtu comparison in Chapter 4).
- **PG&E:** The weighted NTGRs for PG&E declined by 9 percent compared to PY2013 evaluation results but are comparable to PY2010-2012 results.
- **SCE:** Current cycle results have declined by 19 percent compared to PY2013, based on an NTGR of 0.46 for PY2014 projects versus an NTGR of 0.57 for PY2013. PY2014 results are 6 percent lower than PY2010-2012 findings.
- **SDG&E:** NTGRs for SDG&E's projects have declined by 14 percent compared to PY2013 results but have improved by 6 percent compared to PY2010-12 results.
- **SCG:** For SCG the weighted NTGR across all projects is 0.62. This represents a 5 percent decline from the PY2013 average NTGR of 0.66. However, PY2014 results have improved by 28 percent when compared to PY2010-12 results.

Behind the NTGRs calculated for each project are 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 was also found in the 2010-2012 and 2013 evaluations, factors that negatively impact program influence, and therefore reduce the resulting NTGR, include corporate policy, the addition of measures that work directly with existing equipment, measure installations that are consistent with corporate practices, and measures installed to replace old or failing equipment. Factors that positively impact program influence and increase the resulting NTGR include project decision making following discussions with program staff, projects where energy efficiency is a primary goal, and projects where the program incentive represents a high percentage of the project cost.

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 derived by first calculating the product of the ex-post GRRs and the NTGRs, then calculating the same product based on ex-ante GRR and NTGR estimates, and finally taking the ratio of those two terms. The resulting ratio is a multiplier that describes all evaluation adjustments relative to ex-ante savings claims.

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 90 EAR projects were part of the IALC 2014 population: 29 were installed in PG&E territory, 15 in SCE territory, five in SDG&E territory, and 41 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. Nine of these 90 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. Ex-Ante LC Gross Savings	2,973,082,004	519,130	278,958,430
b. Ex-Ante GRR	0.90	0.90	0.91
c. Ex-Ante Adjusted Gross Savings (c = a x b)	2,684,743,653	468,204	252,652,530
d. Ex-Ante NTGR	0.68	0.70	0.68
e. Claimed Net Savings (e = c x d)	1,822,583,515	326,737	172,655,064
f. Ex-Ante GRR x NTGR (f = b x d)	0.61	0.63	0.62
Evaluation			
g. Evaluation LC GRR	0.62	0.74	0.62
h. Evaluated Gross Results (h = a x g)	1,838,317,902	383,759	172,485,749
i. Evaluation NTG Ratio	0.51	0.51	0.51
j. Evaluated Net Results (j = h x i)	930,162,875	194,177	87,275,351
k. Evaluation GRR x NTGR (k = g x i)	0.31	0.37	0.31
l. Evaluated Net Realization Rate (l = j / e)	0.51	0.59	0.51

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. Ex-Ante LC Gross Savings	3,355,045,379	511,946	2,966,020
b. Ex-Ante GRR	0.90	0.90	0.90
c. Ex-Ante Adjusted Gross Savings (c = a x b)	3,032,257,710	461,886	2,670,784
d. Ex-Ante NTGR	0.63	0.63	0.44
e. Claimed Net Savings (e = c x d)	1,910,797,067	290,148	1,176,983
f. Ex-Ante GRR x NTGR (f = b x d)	0.57	0.57	0.40
Evaluation			
g. Evaluation LC GRR	0.58	0.46	0.58
h. Evaluated Gross Results (h = a x g)	1,941,459,957	236,259	1,716,343
i. Evaluation NTG Ratio	0.46	0.46	0.46
j. Evaluated Net Results (j = h x i)	889,114,487	108,198	786,020
k. Evaluation GRR x NTGR (k = g x i)	0.27	0.21	0.27
l. Evaluated Net Realization Rate (l = j / e)	0.47	0.37	0.67

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. Ex-Ante LC Gross Savings	646,031,867	109,858	17,058,472
b. Ex-Ante GRR	0.90	0.90	0.90
c. Ex-Ante Adjusted Gross Savings (c = a x b)	582,400,372	99,078	15,374,730
d. Ex-Ante NTGR	0.62	0.60	0.62
e. Claimed Net Savings (e = c x d)	358,423,754	59,471	9,491,413
f. Ex-Ante GRR x NTGR (f = b x d)	0.55	0.54	0.56
Evaluation			
g. Evaluation LC GRR	0.63	0.63	0.63
h. Evaluated Gross Results (h = a x g)	409,878,100	68,876	10,822,831
i. Evaluation NTG Ratio	0.51	0.51	0.51
j. Evaluated Net Results (j = h x i)	209,600,914	35,222	5,534,512
k. Evaluation GRR x NTGR (k = g x i)	0.32	0.32	0.32
l. Evaluated Net Realization Rate (l = j / e)	0.58	0.59	0.58

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

Impact Element	LC Gas Savings Therms/year
Tracking	
a. Ex-Ante LC Gross Savings	224,927,098
b. Ex-Ante GRR	0.91
c. Ex-Ante Adjusted Gross Savings (c = a x b)	205,217,195
d. Ex-Ante NTGR	0.49
e. Claimed Net Savings (e = c x d)	100,975,736
f. Ex-Ante GRR x NTGR (f = b x d)	0.45
Evaluation	
g. Evaluation LC GRR	0.49
h. Evaluated Gross Results (h = a x g)	109,811,638
i. Evaluation NTG Ratio	0.62
j. Evaluated Net Results (j= h x i)	68,519,951
k. Evaluation GRR x NTGR (k = g x i)	0.30
l. Evaluated Net Realization Rate (l = j / e)	0.68

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, and Program Practices Assessment (PPA) activities. Extensive reporting of findings and recommendations is presented in Chapter 6 of this report. Readers are encouraged to examine Chapter 6 for additional details and context regarding the overarching recommendations outlined below.

At a summary level, the detailed recommendations in this report have been condensed as follows:

- To more accurately estimate ex-ante savings, the PAs should:
  - Improve documentation and reporting of project EUL,[19] including a review of evaluation EUL conclusions/rationale in an effort to improve EUL estimates and LC GRR results;

[19] It is notable that the evaluation estimate of EUL differed from the PAs estimate 48 percent of the time. For those instances the evaluation-derived average EUL was smaller than the ex-ante average EUL by over four years, representing a 26 percent reduction in the ex-ante EUL estimate for that subset of observations. As noted in Chapter 3, with the exception of PG&E LC GRR results were lower than FY GRR results and this EUL difference was *a key factor* driving down the LC GRR results.

  - Improve quality control of project operating conditions verification and normalization, 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. SCG consistently documents and reports on key project characteristics and conclusions in the project application files, and SCG was also successful in transferring documented conditions from the application forms and reports into the tracking system in a manner that is consistent with CPUC guidance and decisions. PG&E also performed relatively well in terms of documentation and tracking.
- 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. While the need for PA attention to each recommendation varies based on the results of this evaluation, in general all recommendations apply to all PAs to some degree. The general recommendations provided below should be addressed by all PAs. PA-specific recommendations are detailed in Chapter 6.

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

Key Recommendations by Topic Area
Clearly identify project event in terms of natural replacement, replace on burnout, early replacement, new construction, add-on equipment, and system optimization, and set the appropriate baseline accordingly
Disseminate information on baseline selection to ensure best practices across program staff, implementers and customers
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
Where applicable, the PAs need to carefully investigate and document the age, condition and functionality of existing equipment and operations, and use these to establish proper baselines
When baseline conditions are defined by the pre-existing systems the PAs should utilize measured data to define those conditions, and thoroughly document the pre-existing conditions for the purposes of establishing baseline
To improve project eligibility screening it is recommended that the PAs ensure that incented measures exceed the ISP / code baseline
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
The PAs should calibrate models and true-up savings claims based upon post-installation data, such as equipment usage profiles, equipment specifications and model inputs
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 eligibility 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 22 Final Site Reports (FSRs) listed in Table 3-6 to identify the specific reasons that led zero or negative savings, and use those lessons learned to improve related project practices

Key Recommendations by Topic Area
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, implementing and overseeing CPUC guidance should be increased
Better ex-ante documentation is needed supporting project cost-effectiveness and lifecycle saving estimation
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 known free riders by upselling to higher efficiency levels, multi-measure solutions and continuous energy improvement
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 outreach channels to engage individual customers and encourage a long-term energy efficiency-based focus

2

Introduction and Background

This report presents draft results from the impact evaluation of the 2013-2014 California Program Administrator's (PAs)[20] energy efficiency programs, focusing on nonresidential custom measures. This effort is managed by the CPUC Energy Division (ED) staff and is referenced as the Industrial, Agricultural and Large Commercial (IALC) Roadmap on the CPUC ED public documents website.[21] The *IALC Research Plan* dated July 2014 and the *IALC Research Plan Addendum* dated June, 2015, provide additional detail on the evaluation effort; these evaluation plans are available on the ED public documents website.[22] Readers may also want to familiarize themselves with a number of other relevant CPUC sources that are referenced throughout this report.[23] This includes a nonresidential Net-to-Gross (NTG) methods document,[24] which can also be found on the ED public documents website.[25] 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. 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.

---

[20] 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.

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

[22] http://www.energydataweb.com/cpuc/deliverableView.aspx?did=1133&uid=0&tid=0&cid=
http://www.energydataweb.com/cpucFiles/pdaDocs/1307/IALC%20Research%20Plan%20AddendumPY2014_June2015_Final.pdf

[23] 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.

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

[25] 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[26] 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 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 and net-to-gross ratios for custom projects across programs.[27]

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. This IALC impact report addresses findings for the 2014 claim period; the IALC 2013 report is available on the ED public documents website.[28]

The CPUC organized all of its consultant evaluation and research work for PY2013-2015 into roadmaps.[29] 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,

[26] In 2014 PG&E had 58 programs that include custom projects, SCE had 46 such programs, SDG&E had 6, and SCG had 5.

[27] Results in this evaluation are developed by PA and project size strata. Realization rate results are applied to all projects in a given PA/stratum because, during sampling, all projects, regardless of program, have an equal chance of selection by PA/stratum.

[28] http://www.energydataweb.com/cpucFiles/pdaDocs/1341/IALC%202013%20Report%20Final%20071715.pdf

[29] See the 2013-2015 Energy Division & Program Administrator Energy Efficiency Evaluation, Measurement and Verification Plan Version 5, available at www.energydataweb.com under the link for Energy Efficiency EM&V Plans.

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.[30]

Energy savings claims from the measures assigned to the IALC roadmap represent a significant contribution to the overall savings portfolios for the PAs' 2014 energy efficiency programs, accounting for about 24 percent of statewide electric savings claims and 61 percent of statewide positive gas savings claims. In 2014 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 477 GWh and 74 MW. Statewide PA positive gas savings claims for measures assigned to the IALC roadmap total 33 million Therms.

California PA-led custom programs are currently evaluated on an annual basis. The final version of the data necessary for the 2014 IALC evaluation was received by October 29th, 2015. The majority of the evaluation field work and data collection activities took place in the second half of 2015, and the analysis and reporting was completed in the first half of 2016. Evaluation of the 2015 programs is beginning in in the first half of 2016. This annual evaluation schedule results in faster feedback to the PAs with regard to their program activities and supports the Efficiency Savings and Performance Incentive (ESPI) award.[31]

The most recent PA data extract, which reflects cumulative PA program activity through the fourth quarter of 2014,[32] was used to summarize the 2014 claimed energy savings associated with the PA portfolios, as well as the savings assigned to the IALC roadmap's custom impact.[33] These savings are reported in Table 2-1.

---

30 Nonresidential custom lighting projects are evaluated under the Commercial Roadmap and codes and standards are evaluated under the Codes and Standards Roadmap.

31 CPUC Decision 13-09-023 established the ESPI mechanism, which awards PAs for performance in both resource and non-resource activities supporting energy efficiency.

32 Savings in the Q4, 2014 database were frozen as of October 29th, 2015.

33 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.

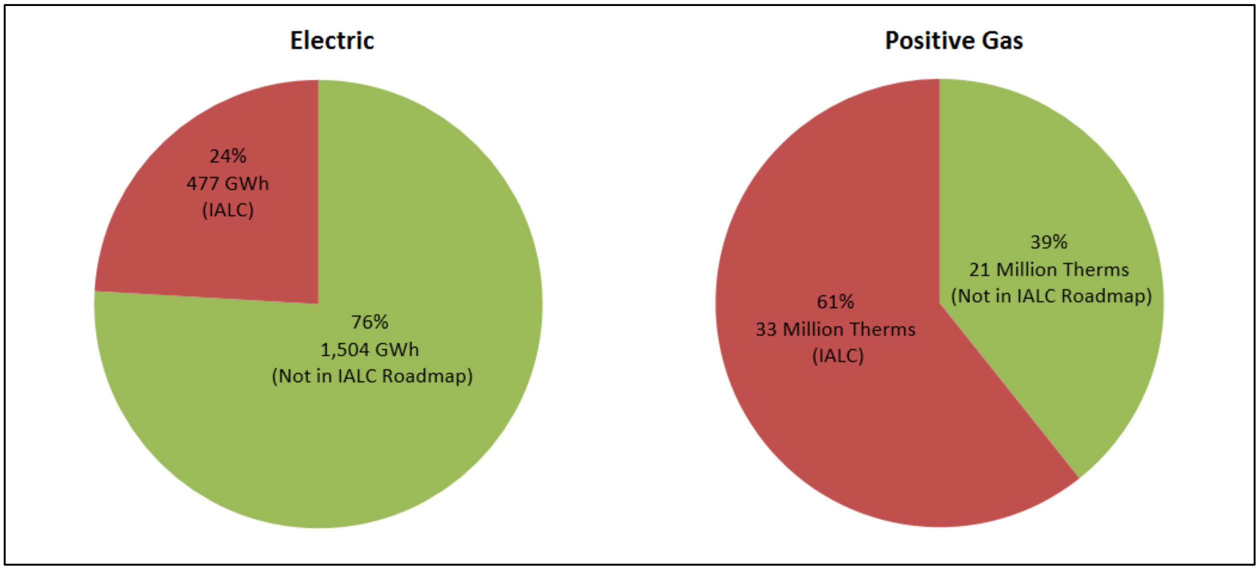
Table 2-1: 2014 Claimed Energy Impacts by PA for Projects in the IALC Roadmap, and in the Portfolio

Claimed Impacts by PA			
PA	**Electric Energy (GWh)**	**Electric Demand (MW)**	**Gas Energy (Million Therms)**[34]
2014 PA Savings Claims, IALC Roadmap			
PG&E	198	32	18
SCE	237	35	0
SCG	-	-	14
SDG&E	42	7	1
Total	477	74	33
Total 2014 PA Savings Claims			
PG&E	800	156	30
SCE	996	172	1
SCG	12	4	21
SDG&E	174	34	3
Total	1,981	366	55
IALC Percentage of Total PA Savings Claims			
PG&E	25%	21%	60%
SCE	24%	21%	19%
SCG	0%	0%	65%
SDG&E	24%	20%	55%
Total	24%	20%	61%

Figure 2-1 contrasts the IALC roadmap savings claims with total portfolio claims for 2014.

[34] 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: 2014 IALC Roadmap Savings Claims Relative to Portfolio Claims



In 2013, the IALC evaluation was further divided into two Work Orders in an effort to isolate and report separately on whole building new construction projects. However, for PY2014 these two impact evaluations were combined under the custom impact evaluation. In general, under the custom evaluation, NRNC whole-building projects are treated like any other points in the program population or evaluation sample. That is, NRNC whole-building project results are pooled with other projects to derive aggregate impact results, including gross realization rates (GRRs) and net-to-gross ratios (NTGRs).

During the process of identifying the IALC claims as Custom and NRNC, two additional groups 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 2014 evaluation. The 4,337 records corresponding to SCE pump testing that were assigned to the IALC roadmap in 2014 were not evaluated. Additionally, there were 103 HVAC records that were assigned to the IALC roadmap after 2014 field work was completed and were therefore not evaluated.[35] Table 2-2 shows how the records assigned to the IALC roadmap in 2014

[35] There were an additional 8,300 HVAC records from PG&E that were also assigned to the IALC roadmap after 2014 fieldwork was completed. These records correspond either to HVAC diagnostic or maintenance measures, or to "contract signing" measures (program PGE21015). These claims receive incentives but claim zero savings and were also not evaluated. All measures claiming savings under the IALC roadmap that were not evaluated were "passed through" (i.e. claimed savings were accepted as-is and were not subjected to the application of evaluation-estimated gross realization rates).

were further separated into the Pump Testing, HVAC, NRNC, and Custom groups. Only positive gas claims are shown.

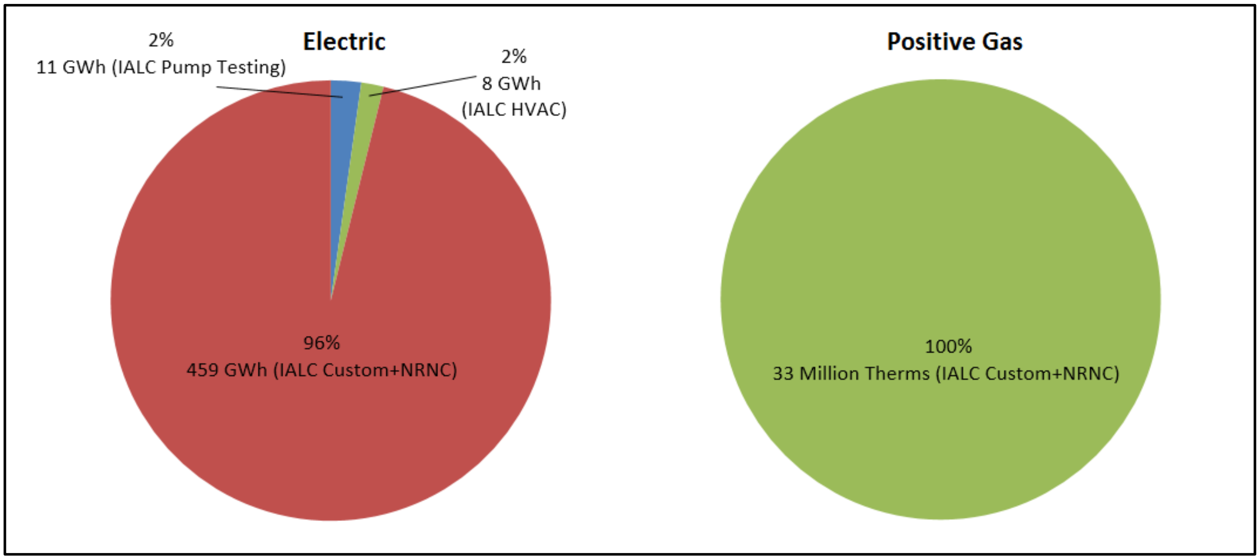
Table 2-2: 2014 Claimed Energy Impacts for the IALC Roadmap, by Group and PA

Claimed Impacts by PA			
PA	**Electric Energy (GWh)**	**Electric Demand (MW)**	**Positive Gas Energy (Million Therms)**
IALC 2014 PA Savings Claims			
PG&E	198	32	18
SCE	237	35	0
SCG	0	0	14
SDG&E	42	7	1
Total*	477	74	33
IALC 2014 Pump Testing PA Savings Claims - Not Evaluated			
PG&E	0	0	0
SCE	11	2	0
SCG	0	0	0
SDG&E	0	0	0
Total*	11	2	0
IALC 2014 HVAC PA Savings Claims - Not Evaluated			
PG&E	4	0	0
SCE	3	0	0
SCG	0	0	0
SDG&E	0	0	0
Total*	8	1	0
IALC 2014 Whole Building NRNC PA Savings Claims - Evaluated			
PG&E	41	12	3
SCE	19	5	0
SCG	0	0	0
SDG&E	11	3	0
Total*	70	20	3
IALC 2014 Custom Savings Claims - Evaluated			
PG&E	153	20	15
SCE	204	28	0
SCG	0	0	14
SDG&E	31	3	1
Total*	388	51	30

* Totals may not sum due to rounding.

Figure 2-2 presents the Custom and NRNC portions of the IALC roadmap ex-ante savings claims that were included in the 2014 evaluation. The figure also shows the HVAC and Pump Testing portions that are not in scope in 2014. Only positive gas claims are shown.

Figure 2-2: 2014 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.[36] 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 estimates. That is, unless otherwise noted, the remainder of the report refers to ex-ante savings that are calculated as PA savings estimates from which the 90 percent default PA RR is backed out (i.e., PA savings claim / 0.9 = gross unadjusted savings estimate). In addition, as specified in the *IALC Research Plan*, only positive unadjusted savings estimates were evaluated (refer to Table 2-3 for totals).

36 The EAR process involves an M&V-level of review for PA projects that are under development, prior to installation and subsequent savings claims by the PAs. 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).

Evidence suggests that some projects that have been "touched" by the EAR process do not always have PA RR = 1.0 in the tracking data. For any point in the evaluation M&V sample, the evaluation conducts a careful examination of whether or not that point has been part of the EAR process, and instances have been identified where PA RR is set equal to 0.90 but evidence suggests some level of involvement in EAR. A key issue is that the evaluation needs to both identify EAR points in the custom sample frame and identify those that have been substantially influenced (for example, based on "frozen" status, but not those that were selected and subsequently "waived"). Since the PAs are instructed by the CPUC to ID such EAR points using PA RR equal to 1.0, the evaluation uses this tracking data-based source to identify EAR points (PA RR=1.0).

Table 2-3: 2014 Unadjusted Positive Gross Savings Estimates Evaluated by IALC Custom, by PA

PA	Positive Electric Energy (GWh)	Positive Electric Demand (MW)	Positive Gas Energy (Million Therms)
IALC 2014 Evaluated Unadjusted Positive Gross Savings			
PG&E	215	36	20
SCE	247	37	0
SCG	0	0	15
SDG&E	48	8	2
Total	509	80	36

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.[37]
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

[37] 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

practices.[38] 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.

In order to more fully answer these researchable questions, the Custom evaluation collected information from 150 gross impact points and 196 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 evaluation. 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 or after (i.e. those projects expected to be influenced by EAR recommendations) versus all earlier applications.[39] 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 and 2013 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 2014 study. Since the IALC Custom evaluation was expected to provide results at the PA-level, M&V and NTG samples were designed and implemented at the PA-level.

---

[38] See NR-5 Nonresidential Best Practices Report at http://www.eebestpractices.com/pdf/BP_NR5.PDF

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

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.[40] 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 Therm ex-post savings for each PA.

2.3 Structure of the Report

Table 2-4 shows the overall organizational structure of this report. Although overarching findings and recommendations are 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	**Gross 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
6	**Detailed Findings and Recommendations**	Improvement to program gross and net impact performance is examined, based on findings and recommendations that stem from the evaluation results

[40] 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.

3

Gross Impact Results

This chapter presents quantitative and qualitative gross impact results for the 2014 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)[41] domain and size stratification. Results are also presented for energy metrics – source energy (MMBtu)[42] and electric demand (kW).

Stratified sampling was implemented for custom measures installed in 2014 by each PA separately: 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 PA's lifecycle ex-ante estimate of impacts.[43]

[41] 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.

[42] 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.

[43] 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 combination 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 and ISP applicability
- Evaluation Guidance for Site Specific Analysis, update dated September 18, 2014[44]
- Approved List of ISP Studies[45]

3.1 Project-Specific Gross Impact Summary

Gross impact evaluation results are supported by 150 M&V sample points. A sample point is defined as one or more tracking system records representing measures that were installed at the same site under the same ProjectID or ApplicationCode. These 150 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 2014 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 150 main gross M&V points targeted for completion. However, given customer willingness to participate and other factors, the final gross sample consisted of 145 main points and 5 backup points, collectively representing 36% of ex-ante MMBtu savings estimates 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 Disposition by PA

PA	Completed M&V Points (n)		Percent of Ex-Ante MMBtu Savings Estimates
	Main	Backup	M&V Sample
PG&E	43	0	31%
SCE	39	3	30%
SDG&E	34	1	49%
SCG	29	1	55%
All PAs	**150**		**36%**

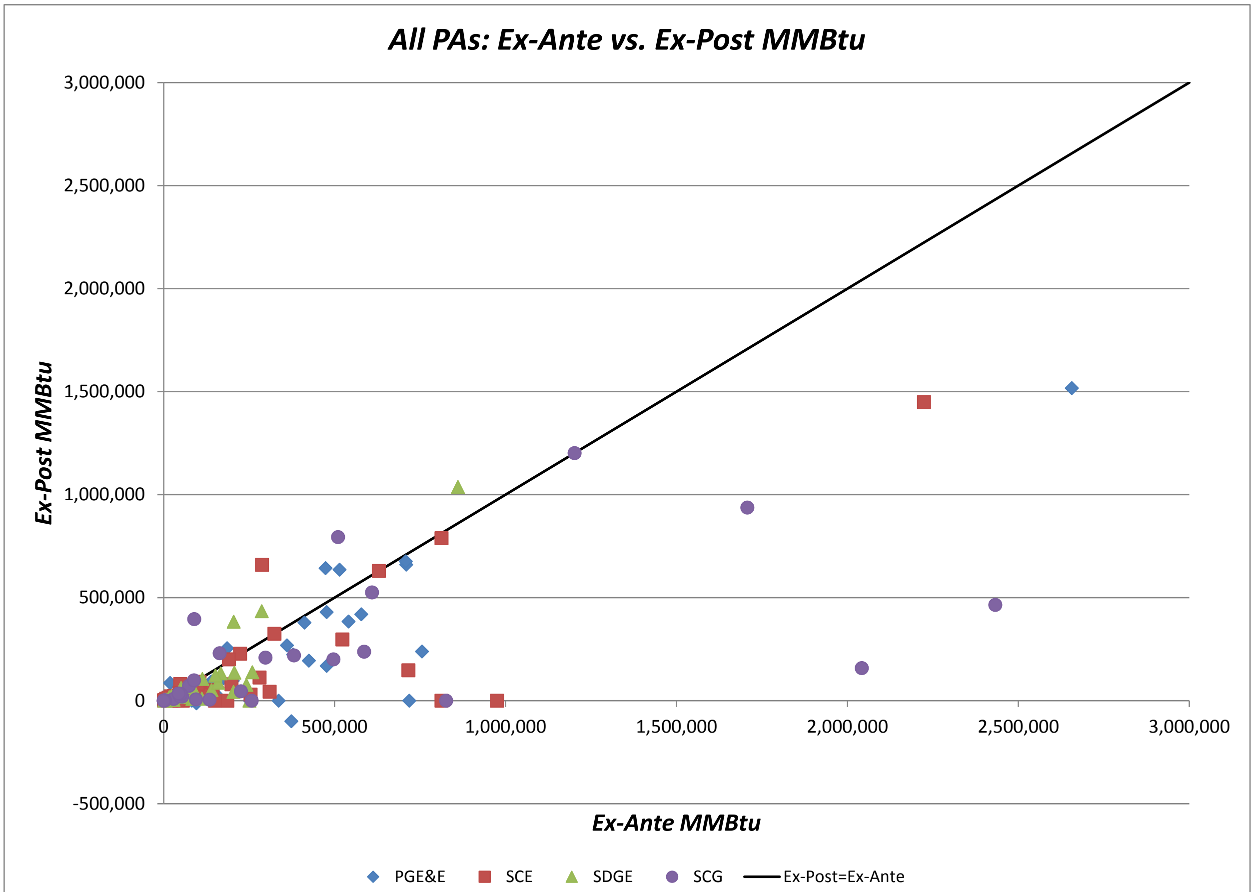
[44] Industrial, Agricultural and Large Commercial Evaluation Guidance available at www.energydataweb.com/cpuc/. Select the search tab, and from the drop down menus, select Portfolio Cycle 2013-2014 and Work Order (ED_I_IAL_2-Itron) 1314 IALC Impact. Direct link: http://energydataweb.com/cpucFiles/pdaDocs/1256/Evaluation%20Guidance%20Questions%20for%20Site%20Specific%20Analysis_2014_0918.pdf

[45] http://www.cpuc.ca.gov/General.aspx?id=4133

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[46] 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). Of the 150 M&V points, 149 projects are included in the figure.[47] PA-specific plots are included in Appendix B.

Most of the sampled projects yielded lifecycle GRRs that fall below the unity line in the chart.

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



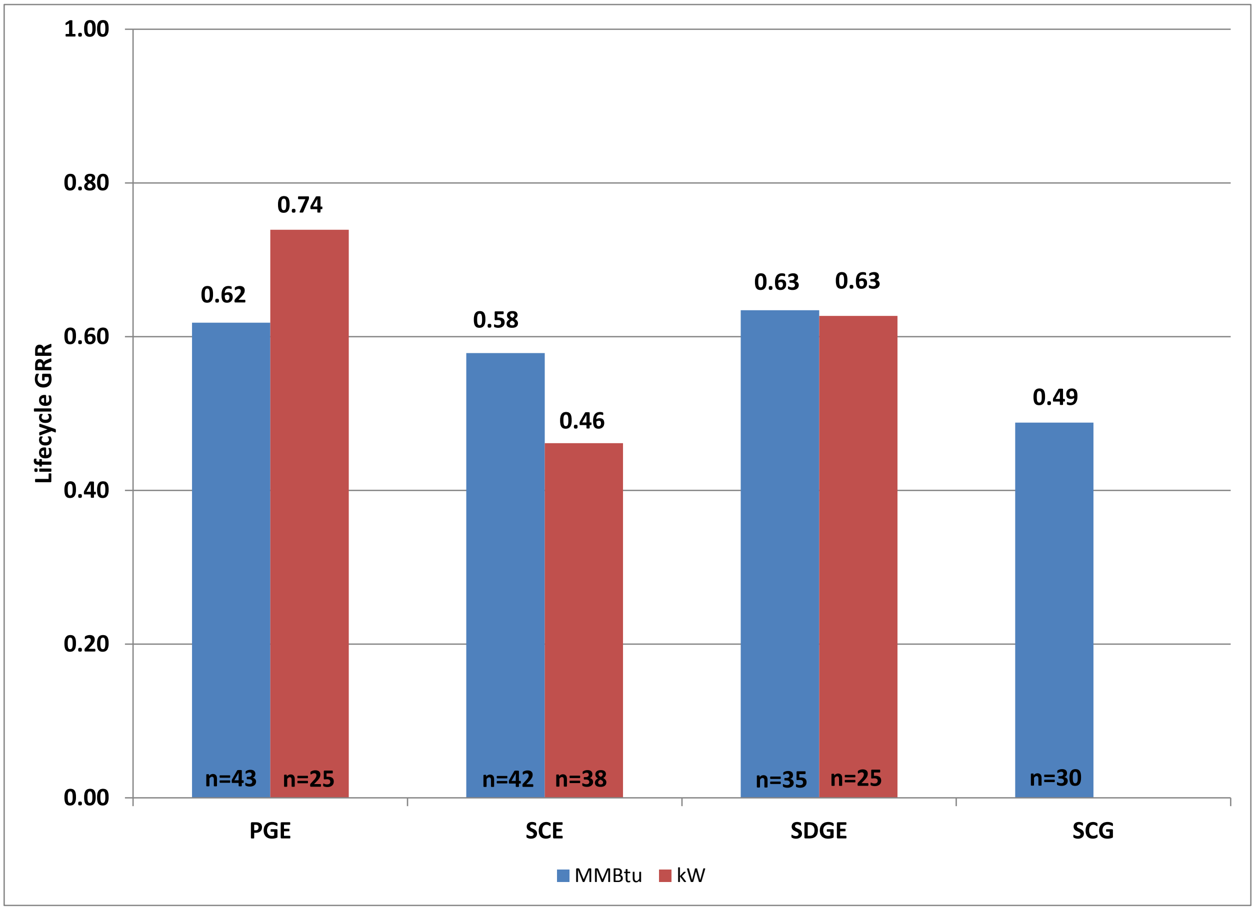
[46] This figure compares "engineering estimates" for both ex-ante MMBtu and ex-post MMBtu. That is, if the PA-claimed ex-ante savings for a record includes the PA RR=0.9 adjustment, that adjustment was removed for the purpose of this comparison.

[47] No extreme lifecycle GRRs (very high or very low) were found in the 2014 evaluation; this report therefore does not include discussions of results generated with- and without extreme points as in the 2010-2012 evaluation. However, one PG&E project is excluded from this figure due to extremely high savings (~5,000,000 ex-ante MMBtu, ~6,500,000 ex-post MMBtu). Results for this project can be seen in the PA specific plots included in Appendix B.

3.2 PA Gross Realization Rate Results

Weighted gross realization rates by PA and energy metric (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 are generally 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.[48] 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	Population Count	Sample Count	% of Savings Sampled	LC Mean GRR	Error Ratio	90% Confidence Interval	LC GRR >1.5	LC GRR =0	LC GRR <0	FY Mean GRR
PG&E ŧ										
MMBtu*	1,244	43	31%	0.62	0.75	0.50 to 0.73	2	6	3	0.59
kW	942	25	14%	0.74	1.67	0.34 to 1.14	5	3	4	0.69
SCE										
MMBtu*	1,161	42	30%	0.58	0.94	0.44 to 0.71	3	9	0	0.64
kW	1,069	38	25%	0.46	1.00	0.34 to 0.58	1	10	1	0.50
SDGE										
MMBtu*	203	35	49%	0.63	0.43	0.57 to 0.70	2	2	0	0.73
kW	139	25	38%	0.63	0.45	0.54 to 0.71	0	2	2	0.67
SCG										
MMBtu	236	30	55%	0.49	0.96	0.36 to 0.62	2	2	0	0.58

* Primary sample was designed and selected based on ex-ante MMBtu savings estimates. Note that the MMBtu and kW sample and population counts are not equal, as not every project included a kW saving claim.

ŧ Unlike the other PAs, PG&E LC GRR results are higher than FY GRR results. This is caused by five PG&E projects with significant weights in their respective strata that were assigned a higher EUL by the evaluation. This caused the LC GRR of those strata to increase compared to the FY GRR, and ultimately caused a slight increase in the weighted LC GRR as compared to the weighted FY GRR.

The error ratio for most domains is similar to the error ratios obtained in the *2010-2012 WO033 Custom Impact Evaluation Final Report* and the *2013 IALC Custom Impact Evaluation Final Report.*[49] In the 2014 sample, the relatively high standard deviation is indicative of the variability in the data rather than stemming from a small sample size. The underlying sample has individual gross realization rates that are widely dispersed between zero (or less) and values exceeding 1.0, which will always result in a large standard deviation regardless of the number of projects sampled. For example, the M&V sample supporting PG&E's kW GRR consisted of 25 projects, five of which had LC GRRs of greater than 1.5 and seven projects with LC GRRs of zero or less, resulting in an error ratio of 1.67.

The mean lifecycle realization rates by PA and energy metric are less than 0.70 for all but one energy metric (the GRR for PG&E kW is 0.74) and are similar to those from the 2010-2012 and

48 The error ratio is a measure of the statistical variation in the gross realization rates.

49 http://www.calmac.org/publications/2010-12_WO033_Custom_Impact_Eval_Report_Final.pdf
http://www.calmac.org/publications/IALC_2013_Report_Final_071715.pdf

2013 evaluations. As a comparison to 2014 results presented in the table above, Table 3-3, Figure 3-3 and Figure 3-4 include LC GRR evaluation results from the 2010-12 and 2013 cycles on a combined MMBtu basis.[50] LC GRR results appear in the upper half of Table 3-3. While none of the LCC GRR results (kW and MMBtu) are statistically different at the 90% confidence level between the three evaluation cycles, 2014 GRR results show an increase for PG&E kW and SDG&E MMBtu and a decrease for SCE kW, SDG&E kW, and SCG MMBtu.

Table 3-3 also includes a similar comparison of FY GRR results in the lower half of the table. Relative to the 2013 custom impact evaluation results PG&E, SDG&E, and SCG FY MMBtu GRRs decreased by about 20 percent, three percent, and 16 percent, respectively. FY GRR results increased for SCE by about 17 percent. It is notable that FY GRRs are an indication of PA performance in conducting ex-ante engineering-based savings estimates and associated PA processes, such as the correct establishment of project type and project baseline, appropriate calculation methods, and proper accounting of operating conditions. LC GRRs are an indication of PA performance for all FY engineering elements plus EUL and early retirement (ER) treatment (including associated RUL and EUL considerations).

[50] 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 and 2014 results.

Table 3-3: 2010-2012, 2013, and 2014 Weighted Project Realization Rates by PA and Energy Metric (MMBtu and kW)

Energy Metric	2010-2012 Mean Gross Realization Rate	2010-2012 90% Confidence Interval	2013 Mean Gross Realization Rate	2013 90% Confidence Interval	2014 Mean Gross Realization Rate	2014 90% Confidence Interval
PG&E LC GRR Results						
MMBtu*	0.63	0.57 to 0.69	0.63	0.57 to 0.70	0.62	0.50 to 0.73
kW	0.46	0.35 to 0.58	0.44	0.28 to 0.61	0.74	0.34 to 1.14
SCE LC GRR Results						
MMBtu	0.61	0.51 to 0.71	0.44	0.34 to 0.54	0.58	0.44 to 0.71
kW	0.57	0.47 to 0.67	0.52	0.43 to 0.62	0.46	0.34 to 0.58
SDGE LC GRR Results						
MMBtu*	0.56	0.47 to 0.66	0.49	0.40 to 0.59	0.63	0.57 to 0.70
kW	0.82	0.46 to 1.17	0.76	0.57 to 0.95	0.63	0.54 to 0.71
SCG LC GRR Results						
MMBtu*	0.64	0.54 to 0.75	0.60	0.48 to 0.72	0.49	0.36 to 0.62
PG&E FY GRR Results						
MMBtu*	0.65	0.59 to 0.70	0.74	0.69 to 0.80	0.59	0.49 to 0.70
kW	0.53	0.41 to 0.66	0.54	0.37 to 0.70	0.69	0.28 to 1.10
SCE FY GRR Results						
MMBtu	0.60	0.54 to 0.67	0.54	0.43 to 0.66	0.64	0.49 to 0.78
kW	0.61	0.53 to 0.70	0.64	0.53 to 0.76	0.50	0.37 to 0.64
SDGE FY GRR Results						
MMBtu*	0.43	0.37 to 0.50	0.75	0.66 to 0.84	0.73	0.65 to 0.80
kW	0.84	0.48 to 1.19	1.02	0.88 to 1.17	0.67	0.52 to 0.81
SCG FY GRR Results						
MMBtu*	0.71	0.58 to 0.84	0.69	0.61 to 0.77	0.58	0.45 to 0.71

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

Figure 3-3: Comparison of 2010-12, 2013, and 2014 Weighted MMBtu LC GRR Results

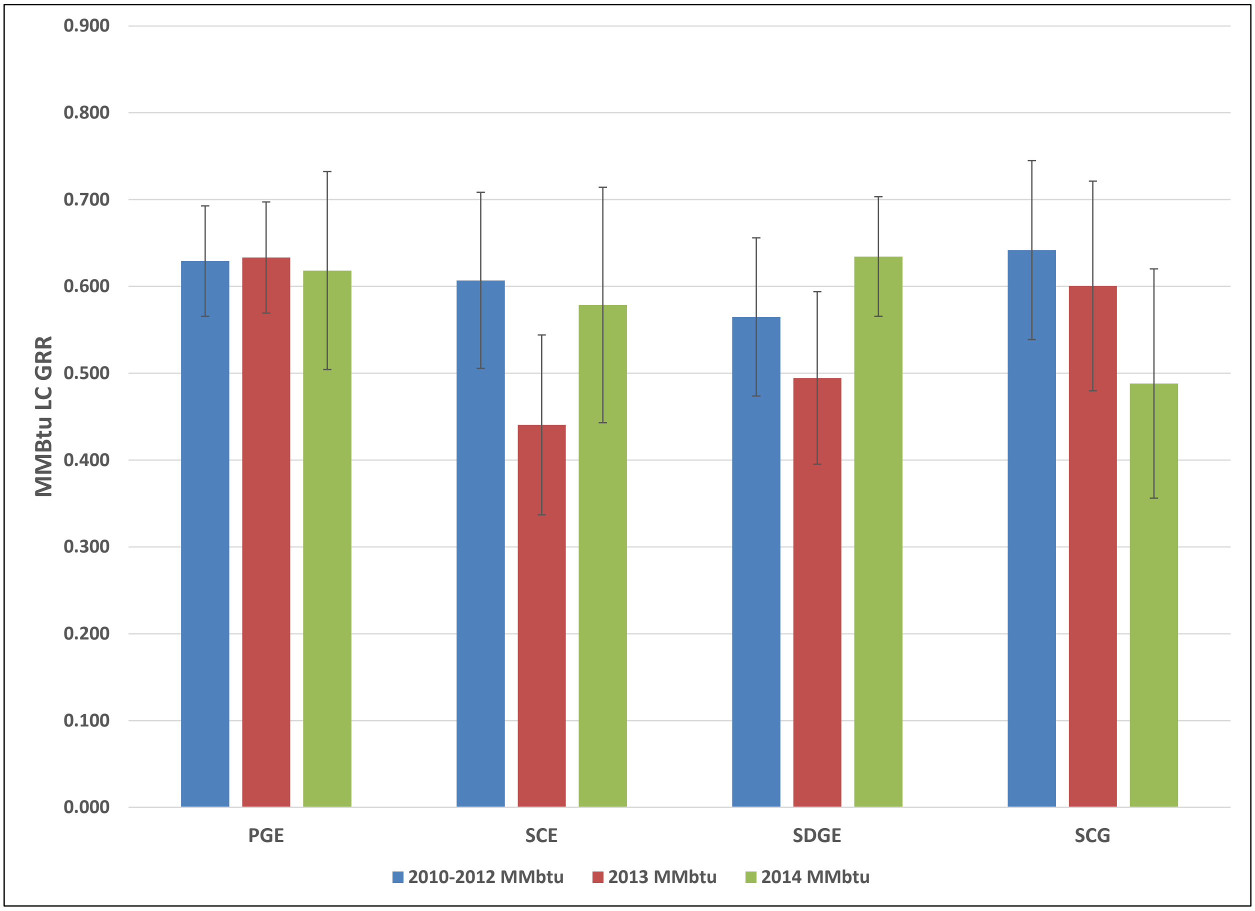
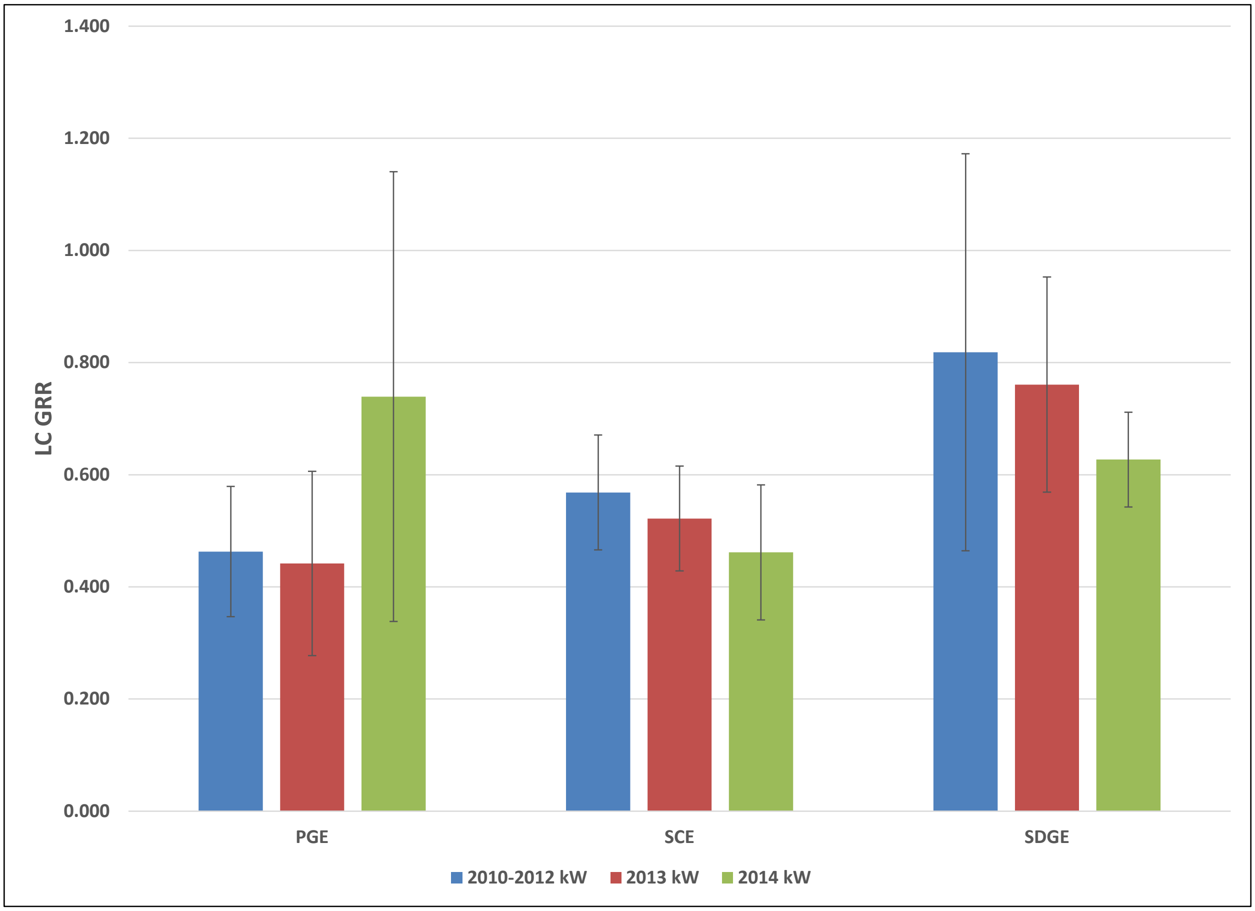


Figure 3-4: Comparison of 2010-12, 2013, and 2014 Weighted kW LC GRR Results



For all PAs except PG&E, weighted lifecycle realization rates in 2014 are lower than the corresponding first year realization rate, which is primarily due to adjustments in measure effective useful life (EUL). In particular, of the 192 measures (some of the 150 projects had multiple measures) evaluated in 2014, the ex-ante EUL was overstated in the tracking system extract for 70 measures (for example: a measure with 5-year life expectancy was assigned a 15-year EUL.) There were also 20 evaluated measures claiming understated EULs, but the upward adjustments for these cases were less significant. Across all PAs and measures in the sample, EULs were overstated by a total of 421 years compared to understatement of 101 years. Section 5.3.5 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 Section 3.4.

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. 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

PA Domain	Strata	Sample Count	% of MMBtu Savings Sampled	Projects with kW Ex-Ante	% of kW Savings Sampled	Weighted LC GRR kW	Weighted LC GRR MMBtu	MMBtu LC GRR > 150%	MMBtu LC GRR = 0%	MMBtu LC GRR < 0%
PG&E	1	2	100%	2	100%	0.60	1.05	0	0	0
	2	11	39%	5	26%	0.92	0.69	0	1	0
	3	10	33%	5	18%	0.43	0.50	0	2	1
	4	10	7%	5	4%	0.59	0.49	1	1	2
	5	10	1%	8	1%	0.94	0.50	1	2	0
SCE	1	8	100%	8	100%	0.44	0.48	0	2	0
	2	9	22%	8	21%	0.72	0.81	1	1	0
	3	9	14%	9	13%	0.39	0.36	1	2	0
	4	8	4%	7	7%	0.09	0.43	0	3	0
	5	8	1%	6	1%	0.59	0.73	1	1	0
SDG&E	1	2	100%	2	100%	1.04	1.28	1	0	0
	2	7	73%	6	49%	0.26	0.67	1	1	0
	3	9	50%	8	42%	0.28	0.42	0	0	0
	4	9	23%	3	9%	1.20	0.70	0	0	0
	5	8	5%	6	5%	0.46	0.31	0	1	0
SCG	1	2	84%	-		-	0.34	0	0	0
	2	5	49%	-		-	0.40	0	1	0
	3	7	89%	-		-	0.82	2	1	0
	4	8	29%	-		-	0.67	0	0	0
	5	8	6%	-		-	0.69	0	0	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. PG&E's stratum 3 and SCE's stratum 4). 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. SDG&E's stratum 1 and SCG's stratum 3).
- The stratum 1 LC MMBtu GRRs for PG&E and SDG&E and stratum 1 LC kW GRR for SDG&E were all greater than 1.0, which served to increase the overall weighted LC GRRs for these PAs.
- The realization rate for kW in the SCE electric domain, stratum 4, is 0.09; this is due to 5 of the 7 sampled projects having GRRs of zero or less than zero.

Given relatively small evaluation sample sizes, individual projects in the sample can have a significant influence over stratum-level results. Additionally, because each stratum has a roughly equal weight in the overall result, these observations illustrate the importance that a small number of projects with low realization rates 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 3-6, these projects with zero or negative savings are due primarily to two factors – baseline selection and lack of eligibility screening. 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 B. This appendix includes anonymized site and record identifiers, the strata, ex-ante savings estimates from the PA tracking systems, gross realization rates, and net to gross ratios.

3.3 EAR Overlap Sensitivity Analysis

The establishment of ex-ante review (EAR) is discussed in CPUC Decision 09-09-047,[51] 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 of and compliance with 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

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

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 90 EAR projects (PA RR=1.0) were part of the IALC 2014 population: 29 were installed in PG&E territory, 15 in SCE territory, five in SDG&E territory, and 41 in SCG territory. The IALC stratified random sampling process selected nine EAR projects for evaluation: two from PG&E, three from SCE, zero from SDG&E, and four from SCG. Table 3-5 shows the first year and the lifecycle MMBtu realization rate results for these nine 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[52] 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.[53]

[52] Since there was no deliberate selection of EAR projects as part of the sample, this sensitivity analysis quantifies the effect of removing these particular EAR points from this particular stratified, random sample. Population-level weighting was also adjusted by removing savings weights for any EAR-reviewed project (PA RR=1.0). Quantifying the effect of removing all EAR points from the entire population of projects was not an objective of the study sample design.

[53] 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 not being EAR-reviewed points.

Evidence suggests that some projects that have been "touched" by the EAR process do not always have PA RR = 1.0 in the tracking data. For any point in the evaluation M&V sample, the evaluation conducts a careful examination of whether or not that point has been part of the EAR process, and instances have been identified where PA RR is set equal to 0.90 but evidence suggests some level of involvement in EAR. A key issue is that the evaluation needs to both identify EAR points in the custom sample frame and identify those that have been substantially influenced (for example, based on "frozen" status, but not those that were selected and subsequently "waived"). Since the PAs are instructed by the CPUC to ID such EAR points using PA RR equal to 1.0, the evaluation uses this tracking data-based source to identify EAR points (PA RR=1.0).

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

PA Domain	Project ID	PA RR	FY GRR-MMBtu	LC GRR-MMBtu
PGE	E40001	1.00	1.09	0.73
	E40511	1.00	0.76	0.71
	Remaining Points (n=41)	0.90	0.58	0.62
	All Sample (n=43)	0.91	0.59	0.62
	Percent change due to EAR points	1%	2%	0%
SCE	F40003	1.00	1.00	1.00
	F40524	1.00	1.03	1.02
	F40626	1.00	1.00	1.00
	Remaining Points (n=39)	0.90	0.61	0.55
	All Sample (n=42)	0.91	0.64	0.58
	Percent change due to EAR points	1%	4%	4%
SDG&E	All Sample (n=35)	0.90	0.73	0.63
	Percent change due to EAR points	-	-	-
SCG	G40503	1.00	1.00	1.00
	G40511	1.00	4.41	4.41
	G40021	1.00	1.05	0.79
	G40578	1.00	0.88	0.88
	Remaining Points (n=26)	0.90	0.44	0.38
	All Sample (n=30)	0.91	0.58	0.49
	Percent change due to EAR points	1%	31%	28%

Observations for Table 3-5:

- Seven of the nine EAR projects sampled have first year GRRs of 1.0 or higher, and the remaining projects have first year GRR of 0.76 and 0.88. For all PAs the weighted first year GRRs of the remaining (non-EAR) projects are lower than the FY GRRs for EAR projects.
- Five of the nine EAR projects have lifecycle GRRs of 1.0 or higher. For two of the remaining four EAR projects (E40001 and G40021) the evaluation found significantly shorter EULs than PA ex-ante EULs specified in the tracking extract.[54] This caused these two lifecycle GRRs to be lower than their first year GRRs. For E40511, PA calculation method errors and observed post installation operating conditions resulted in similar reductions in both FY and LC GRRs. The 12 percent reduction in FY and LC GRRs for

[54] For E40001, the tracking data extract showed an EUL of 15 years while the ex-post EUL was determined to be 10 years. However, the project application file for E40001 correctly states the EUL as 10. This highlights the need for program tracking data to be thoroughly checked and populated with the correct data from the application files.

G40578 were also due to observed operating conditions (a reduction in weekly average hours of operation as compared to EAR sites).

- G40511 had FY and LC GRRs of 4.41. Although this was an EAR point, data from as observed operating conditions resulted in increased savings. Normalizing for increased production also led to significantly higher energy savings.

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 2014 IALC population (and sample) improves the first year weighted GRR by two to 31 percent, and the lifecycle weighted GRR by less than one percent to 28 percent for all PAs.

3.4 Discrepancy Analysis

This section presents an analysis of the discrepancies that account for differences between ex-ante and ex-post savings estimates for the sampled projects. Note that this analysis is based on discrepancies associated with first year MMBtu impacts.[55]

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 categories: operating conditions, calculation method, inappropriate baseline, ineligible measure, inoperable measure, measure count, and tracking database discrepancy.[56] 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:[57]

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

[56] A separate 'Other' category includes less common factors and generally accounts for a relatively small number of projects and percentage change in savings claims. However, for SDG&E the other category accounted for a substantial percentage of change to savings claims.

[57] 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 caused an upward adjustment (where ex-post savings estimates are higher than ex-ante estimates).

- Differences in operating conditions (for example, changes in hours of operation, VSD speeds, return to original operation, changes in production levels, etc.).[58]
- Calculation methods used for ex-post savings estimation were different than those used to estimate ex-ante savings. Some examples include: running whole building SBD project simulations using the non-compliance mode; performing an hourly grid impact assessment where on-site generation is present; 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; and use of billing analyses and interval data, particularly for peak demand savings estimation.
- Inappropriate baseline selection or inappropriate use of baseline conditions for ex-ante savings estimation. Some examples of baseline-related issues are: rejected early replacement claims; new equipment that does not exceed code-, ISP- or regulation-required efficiency levels; and inaccurate baseline or pre-retrofit operating hours.
- 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).

[58] 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.

3.4.1 Summary of Discrepancy Factor Impact

Given multiple tracking records associated with some projects, 192 records associated with the impact sample of 150 projects were examined (3.0 Million MMBtu ex-ante savings). For 17 records, the evaluation found no discrepancies (0.2 MMBtu ex-ante savings were not adjusted). For the balance of 175 records and 2.8 MMBtu ex-ante savings, ex-post estimates were different from ex-ante MMBtu estimates. For some records only downward adjustments were observed, while in others only upward adjustments were observed, and in some instances both downward and upward adjustments were applied. Altogether the downward discrepancies in the sample led to a 40 percent reduction in ex-ante savings estimates, while the upward discrepancies accounted for a 7 percent boost, resulting in a net downward adjustment of 33 percent. 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 corresponds to sampled projects with zero or negative MMBtu GRRs. There were 19 projects for which the ex-post MMBtu impacts were zero, and three 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. Also shown is extent of MMBtu reduction to ex-ante savings estimates for each project, by discrepancy factor.

For projects with zero ex-post MMBtu, the discrepancy factors that occur most frequently are inappropriate baseline and ineligible measure. Note that both of these factors can lead to a zero GRR and that some projects that appear here under the inappropriate baseline heading ultimately led to measure ineligibility. Calculation method was the factor that occurred most frequently for projects with negative ex-post MMBtu.

The evaluation conducted a sensitivity analysis to better understand the influence that these 22 projects had on the resulting weighted LC GRR MMBtu results. Removal of the nine PG&E projects resulted in a 23 percent increase in the LC GRR result, from 0.62 to 0.76. Removal of the nine SCE projects resulted in a 26 percent increase in the LC GRR result, from 0.58 to 0.73. Removal of the two SDG&E projects resulted in a 6 percent increase in the LC GRR result, from 0.66 to 0.69. Removal of the two SCG projects resulted in a 10 percent increase in the LC GRR result, from 0.47 to 0.52.

A sensitivity analysis was also conducted to understand the influence that these 22 projects had on the resulting weighted FY GRR MMBtu results. Removal of the nine PG&E projects resulted in a 26 percent increase in the FY GRR result, from 0.59 to 0.75. Removal of the nine SCE projects resulted in a 28 percent increase in the FY GRR result, from 0.64 to 0.82. Removal of the two SDG&E projects resulted in a 6 percent increase in the FY GRR result, from 0.73 to 0.77. Removal of the two SCG projects resulted in an 11 percent increase in the FY GRR result, from 0.58 to 0.64.

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

ItronID*	First Year MMBtu	First Year MMBtu GRR	Lifecycle MMBtu GRR	Customer Agreement Date	Discrepancy Factor and Related Change to Ex-Ante MMBtu Savings					
					Inapprop. Baseline	Ineligible Measure	Calculation Method	Inoperable Measure	Operating Conditions	Other
E40002	35,934	0.00	0.00	8/1/2012		-9,702	-26,232			
E40004	24,913	-0.27	-0.27	5/31/2013			-31,595			
E40011	11,947	-0.14	-0.14	2/19/2013	-6,441	-4,669			-2,453	
E40524	22,412	0.00	0.00	10/9/2014	-4,258				-18,154	
E40536	16,833	0.00	0.00	2/7/2013		-16,833				
E40587	6,470	0.00	0.00	3/14/2013	-6,470					
E40603	5,505	0.00	0.00	5/19/2014			-5,528			
E41163	148	0.00	0.00	7/3/2014	-148					
E41555	875	0.00	0.00	7/30/2014		-875				
F40001	54,172	0.00	0.00	7/20/2012	-54,172					
F40017	11,262	0.00	0.00	11/2/2012		-11,262				
F40052	2,478	0.00	0.00	4/23/2013	-2,478					
F40054	2,364	0.00	0.00	3/29/2013				-2,364		
F40059	2,100	0.00	0.00	11/14/2011				-2,100		
F40150	668	0.00	0.00	3/4/2013					-668	
F40452	9,381	0.00	0.00	9/6/2012		-9,381				
F40502	65,012	0.00	0.00	3/5/2012		-65,012				
F40525	12,410	0.00	0.00	3/18/2013		-12,410				
G40504	41,307	0.00	0.00	5/30/2013		-41,307				
G40506	25,725	0.00	0.00	10/9/2013				-25,725		
H40048	407	0.00	0.00	9/9/2013		-407				
H40503	16,727	0.00	0.00	1/7/2013						-16,727
Discrepancy Frequency					6	10	3	3	3	1

* None of these projects with zero or negative ex-post savings encountered discrepancies associated with measure counts or recording of information in the tracking system.

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 and upward adjustments by discrepancy factor (including projects with zero and negative ex-post

MMBtu,).[59] This summary includes results for all PAs combined at the top of the table and by individual PA below that. Figure 3-5 displays the same information contained in Table 3-7 for all PAs combined. PA-specific plots can be found in Appendix B.

At the statewide level, downward adjustments affected 148 records and caused a -1.2 MMBtu, or 40 percent reduction, to the 3.0 Million MMBtu ex-ante savings estimate for all sampled projects. Likewise, upward adjustments affected 54 records and caused a 0.2 MMBtu, or 7 percent boost, to ex-ante savings. The net reduction for both upward and downward adjustments is 1.0 MMBtu, or 33 percent. All PA ex-ante savings estimates were more greatly influenced by downward adjustments than upward adjustments.

All factors combined, SCE and SCG savings estimates were most affected by downward adjustments, at 46 percent and 51 percent, respectively, versus 40 percent across all PAs. PG&E and SDG&E have smaller, but still significant, downward savings adjustments of 33 and 31 percent, respectively.

At the statewide level, the discrepancy factors that had the greatest influence on adjustments to ex-ante savings estimates, both in terms of downward adjustments and net change, were operating conditions, calculation method, inappropriate baseline and ineligible measure. This was also true for both PG&E and SCE. For SDG&E the other and tracking discrepancy categories were also influential, and in some cases more influential than the top four factors noted above. For SCG the inoperable measure category was more influential than the calculation method category.

The operating conditions discrepancy factor at the statewide level had the greatest downward effect on ex-post MMBtu savings. This factor led to a 19 percent[60] reduction for downward adjustments alone and a 17 percent reduction on a net change basis. This was among the most important factors for all PAs but impacted SCG the most. For SCG the reduction to savings estimates for this single discrepancy factor, operating conditions, was 37 percent, on both a downward-only and net change basis.

Because the purpose of this discussion is to examine opportunities to improve PA GRR results, this section more thoroughly addresses discrepancy factors that have a downward effect on ex-ante savings estimates. The reader is referred to Table 3-7 for the purposes of a more thorough examination of upward effects by discrepancy factor category.

---

[59] 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. All downward and upward adjustments to ex-ante MMBtu savings estimates are also accounted for in the table, as well as the net change in MMBtu that accounts for both upward and downward adjustments.

[60] This statistic of 19 percent is based on results taken from Table 3-7; 564,520 MMBtu divided by 3,027,607 MMBtu.

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

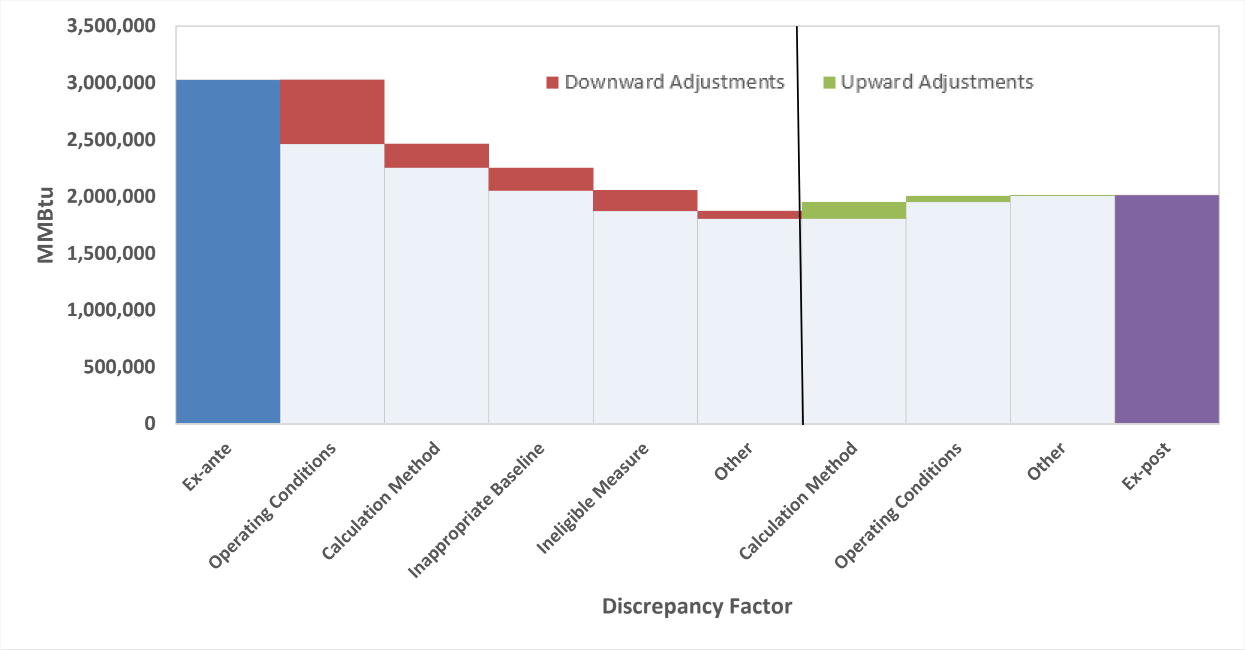


Figure 3-6 shows the distribution of the downward and upward adjustments due to each discrepancy factor across all PAs. Percentages are the fraction of total adjustments (downward and upward, respectively) that are attributed to each discrepancy factor.

Figure 3-6: Distribution of Downward and Upward Adjustments to Ex-ante MMBtu by Discrepancy Factor - All PAs

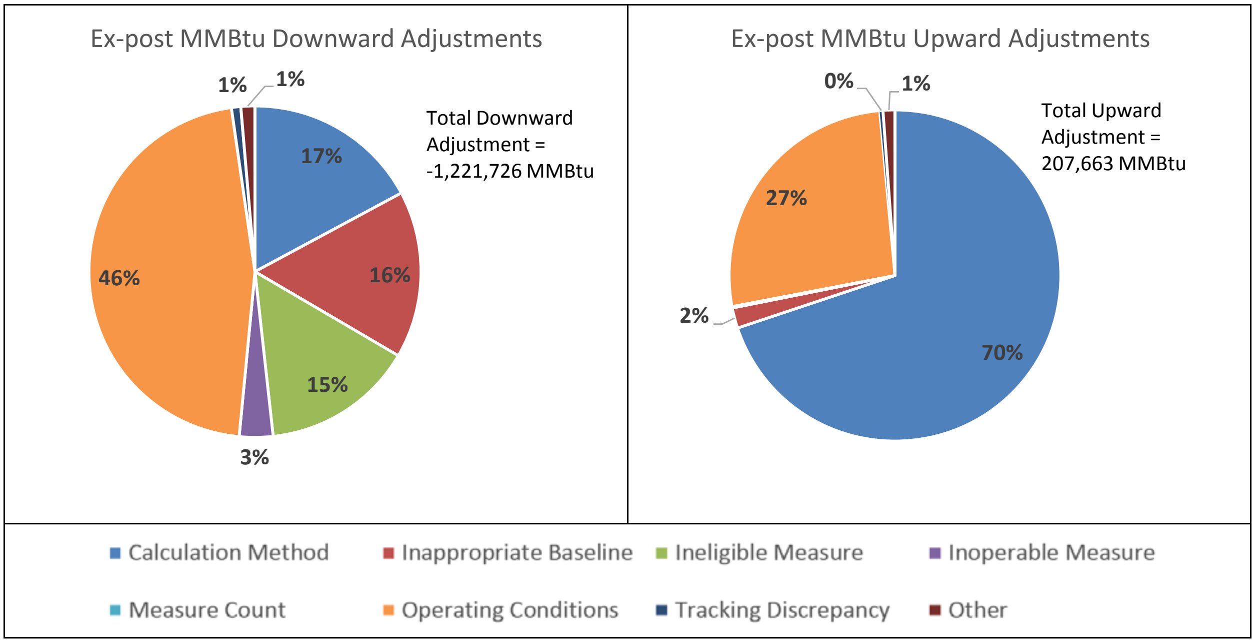


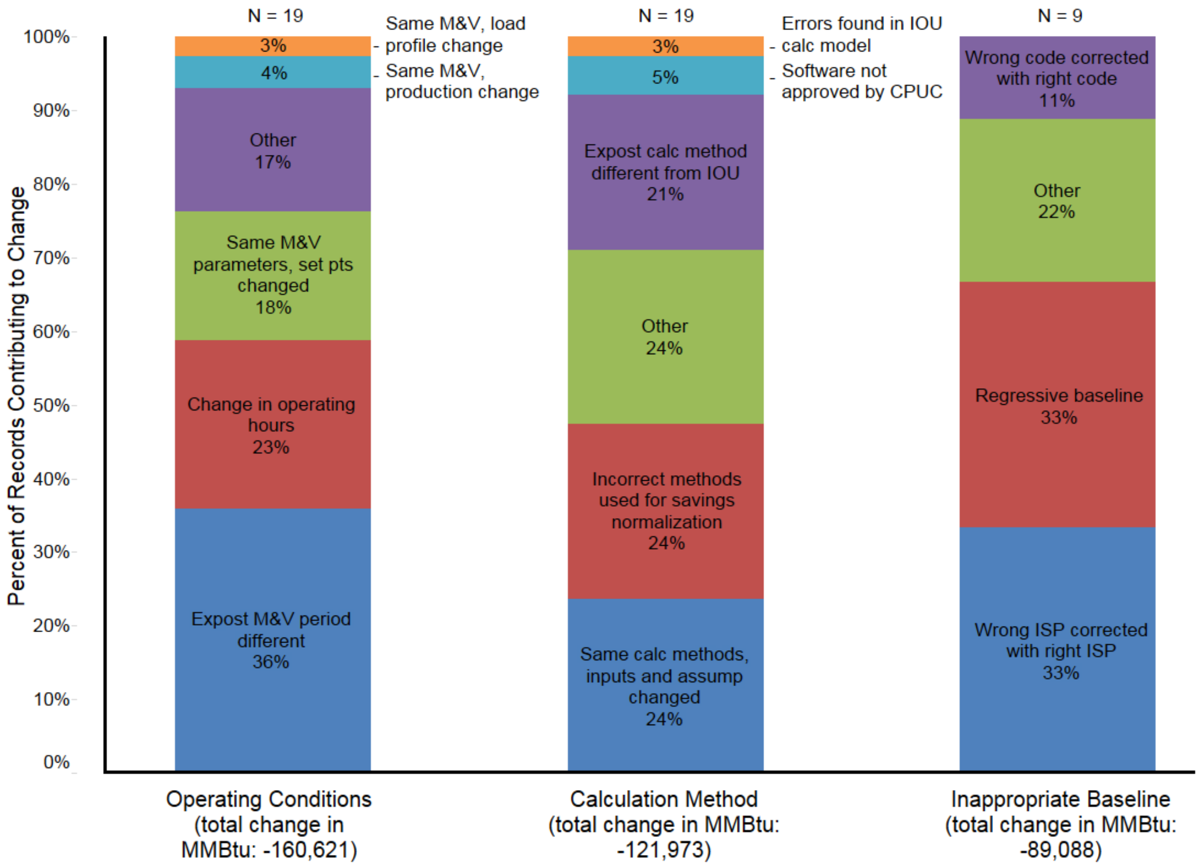
Table 3-7: Records with Ex-Post Downward and Upward Adjustments to Ex-Ante MMBtu Impacts, by Discrepancy Factor; Statewide and by PA

Discrepancy Factor	n Records with Downward Adjustment	n Records with Upward Adjustment	Ex-Ante MMBtu Savings	Ex-Post MMBtu Downward Adjustments	Ex-Post MMBtu Upward Adjustments	Ex-Post MMBtu Net Change
All PAs						
Calculation Method	55	27		-209,422	145,079	-64,343
Inappropriate Baseline	31	2		-199,238	4,141	-195,097
Ineligible Measure	20	0		-180,901	0	-180,901
Inoperable Measure	5	0		-39,891	0	-39,891
Measure Count	1	2		-12	269	257
Operating Conditions	75	27		-564,520	55,097	-509,423
Tracking Discrepancy	7	4		-11,001	700	-10,301
Other	3	2		-16,740	2,377	-14,363
All Records	148	54	3,027,607	-1,221,726	207,663	-1,014,062
PGE						
Calculation Method	19	8		-121,973	17,986	-103,987
Inappropriate Baseline	9	0		-89,088	0	-89,088
Ineligible Measure	5	0		-26,154	0	-26,154
Inoperable Measure	1	0		-9,702	0	-9,702
Measure Count	0	0		0	0	0
Operating Conditions	19	10		-160,621	27,821	-132,799
Tracking Discrepancy	2	0		-2,762	0	-2,762
Other	1	0		-13	0	-13
All PGE Records	40	15	1,233,694	-410,312	45,807	-364,505
SCE						
Calculation Method	13	9		-50,688	32,098	-18,590
Inappropriate Baseline	10	1		-78,148	334	-77,814
Ineligible Measure	6	0		-102,682	0	-102,682
Inoperable Measure	2	0		-4,464	0	-4,464
Measure Count	1	0		-12	0	-12
Operating Conditions	24	8		-87,671	4,555	-83,117
Tracking Discrepancy	3	2		-833	59	-774
Other	0	1		0	266	266
All SCE Records	41	19	711,533	-324,500	37,312	-287,188
SDGE						
Calculation Method	14	6		-21,123	17,794	-3,329
Inappropriate Baseline	6	1		-10,698	3,807	-6,891
Ineligible Measure	7	0		-10,513	0	-10,513
Inoperable Measure	0	0		0	0	0
Measure Count	0	2		0	269	269
Operating Conditions	20	5		-38,069	22,477	-15,592
Tracking Discrepancy	2	2		-7,405	641	-6,765
Other	2	0		-16,727	0	-16,727
All SDGE Records	42	12	333,441	-104,536	44,989	-59,547
SCG						
Calculation Method	9	4		-15,638	77,202	61,563
Inappropriate Baseline	6	0		-21,304	0	-21,304
Ineligible Measure	2	0		-41,552	0	-41,552
Inoperable Measure	2	0		-25,725	0	-25,725
Measure Count	0	0		0	0	0
Operating Conditions	12	4		-278,159	244	-277,915
Tracking Discrepancy	0	0		0	0	0
Other	0	1		0	2,111	2,111
All SCG Records	25	8	748,939	-382,378	79,556	-302,822

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 3-7 addresses the factors that caused the three largest downward adjustments to ex-ante MMBtu for PG&E and provides the frequency (percent of records) with which each sub-category was noted for each of these primary factors.

Figure 3-7: Most Influential Discrepancy Factors that Caused Downward Adjustments for PG&E

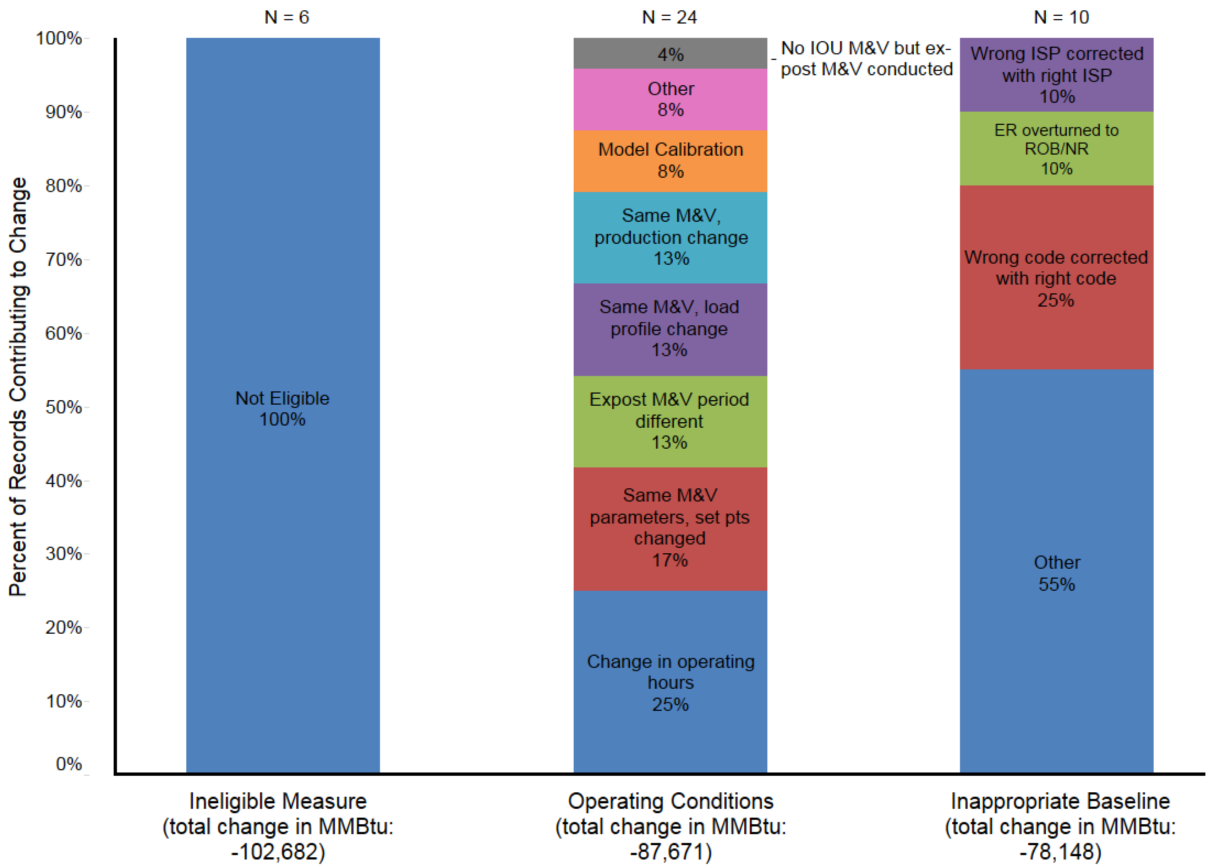


For PG&E the top three discrepancy factors that resulted in a downward adjustment of ex-ante MMBtu impacts were operating conditions, calculation methods, and inappropriate baseline. Changes in operating conditions affected 19 records and resulted in the highest downward adjustment to the ex-ante MMBtu savings estimates, representing a 13 percent overall reduction to the ex-ante MMBtu savings of 1,233,694. Use of different 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 operating hours and set point changes represented the second and third most frequent contributing factors to the reduction of ex-ante MMBtu savings. Production and load profile changes were also noted in a minority of cases.

Calculation method changes during the ex-post analysis resulted in the second highest ex-ante MMBtu savings reduction and represented a 10 percent reduction to the ex-ante MMBtu savings estimate. Nineteen records were affected. The most common sub-categories that explain downward calculation method adjustments were changes to PA inputs and assumptions and improper PA savings normalization. The third most commonly observed sub-category involved the use of a different ex-post savings calculation, followed by the mention of the PA use of unapproved software and modeling errors that were uncovered.

Nine records had downward savings adjustments due to inappropriate baseline issues, resulting in a seven percent reduction to ex-ante MMBtu savings. Three records had incorrectly applied ISP, three records involved regressive baseline requirements, and for one record the evaluation adjusted the project baseline type to a more appropriate code baseline.

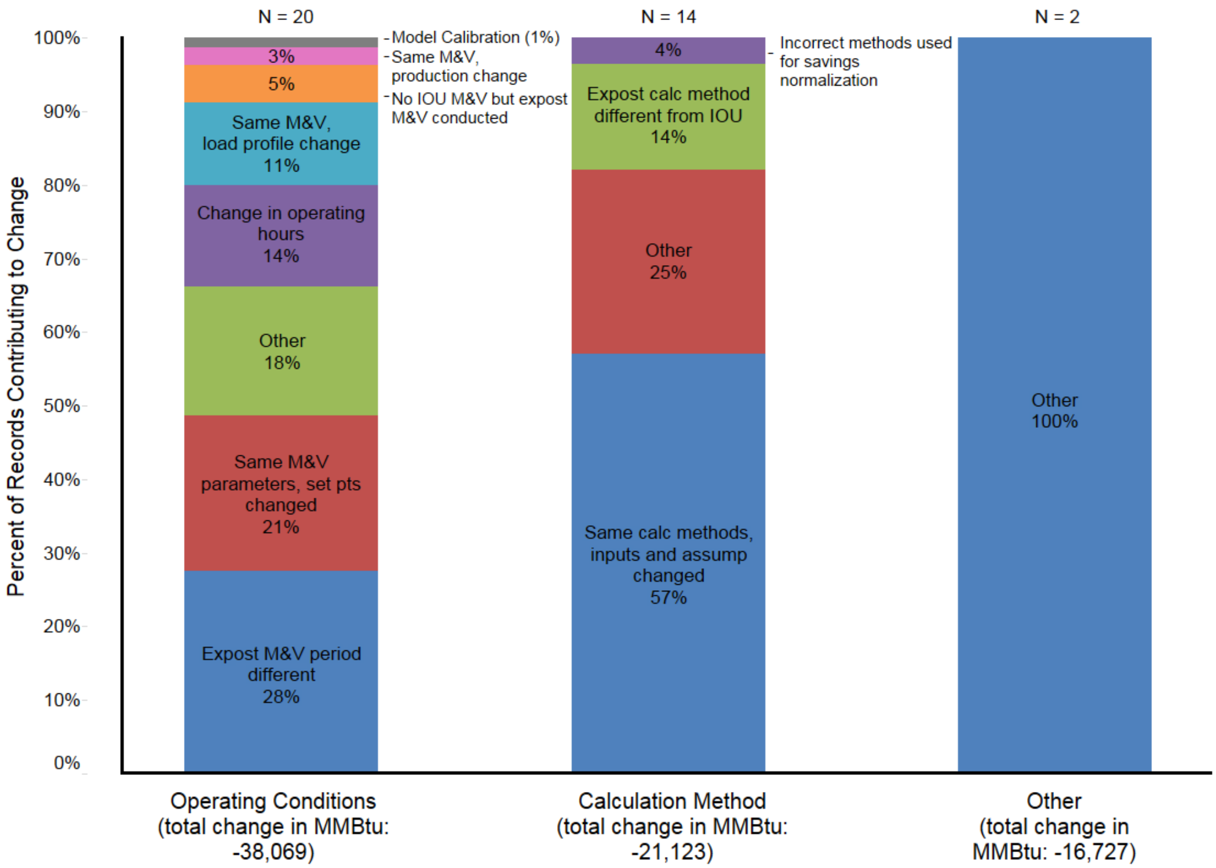
Figure 3-8: Most Influential Discrepancy Factors that Caused Downward Adjustments for SCE



For SCE the top three discrepancy factors that resulted in a downward adjustment of ex-ante MMBtu savings were ineligible measures, operating conditions and inappropriate baseline (Figure 3-8). The six records affected by eligibility resulted in a fourteen percent downward adjustment to ex-ante MMBtu savings estimates of 711,533. A range of issues led to ineligibility, including lack of adequate documentation of pre-installation conditions (and measure rejection due to like-for-like replacement rules), ex-ante models that did not represent site-specific conditions, measure that constituted an operational change, and due to measures that represent mandatory code requirements.

Changes in operating conditions affected 24 records and resulted in the second highest downward adjustment to the ex-ante MMBtu savings estimates, representing a 12 percent overall reduction to the ex-ante MMBtu savings. Changes in operating hours by the evaluation was the most frequently observed sub-category within the operating conditions discrepancy group, while set point changes, use of a different M&V period, load profile changes and production changes were also common categorical explanations used, followed by model calibration and use of ex-post M&V. Ten records were adjusted downwards due to ex-ante use of an inappropriate baseline, resulting in an eleven percent reduction in ex-ante savings estimates. The most commonly noted sub-category was “other” and included further explanation by the assigned engineer, which indicated a mixture of situations, including instances in which the installed equipment was the only technically available solution, where inappropriate set points were included in the PA baseline model, where existing conditions were modeled for a ROB project, where the baseline model incorrectly specified pump head, and where multiple baseline issues were identified for a single record. Additional sub-categorical explanations included evaluation adjustment of the project baseline type to a more appropriate code baseline, incorrectly applied ISP, and ER being overturned to ROB/NR.

Figure 3-9: Most Influential Discrepancy Factors that Caused Downward Adjustments for SDG&E

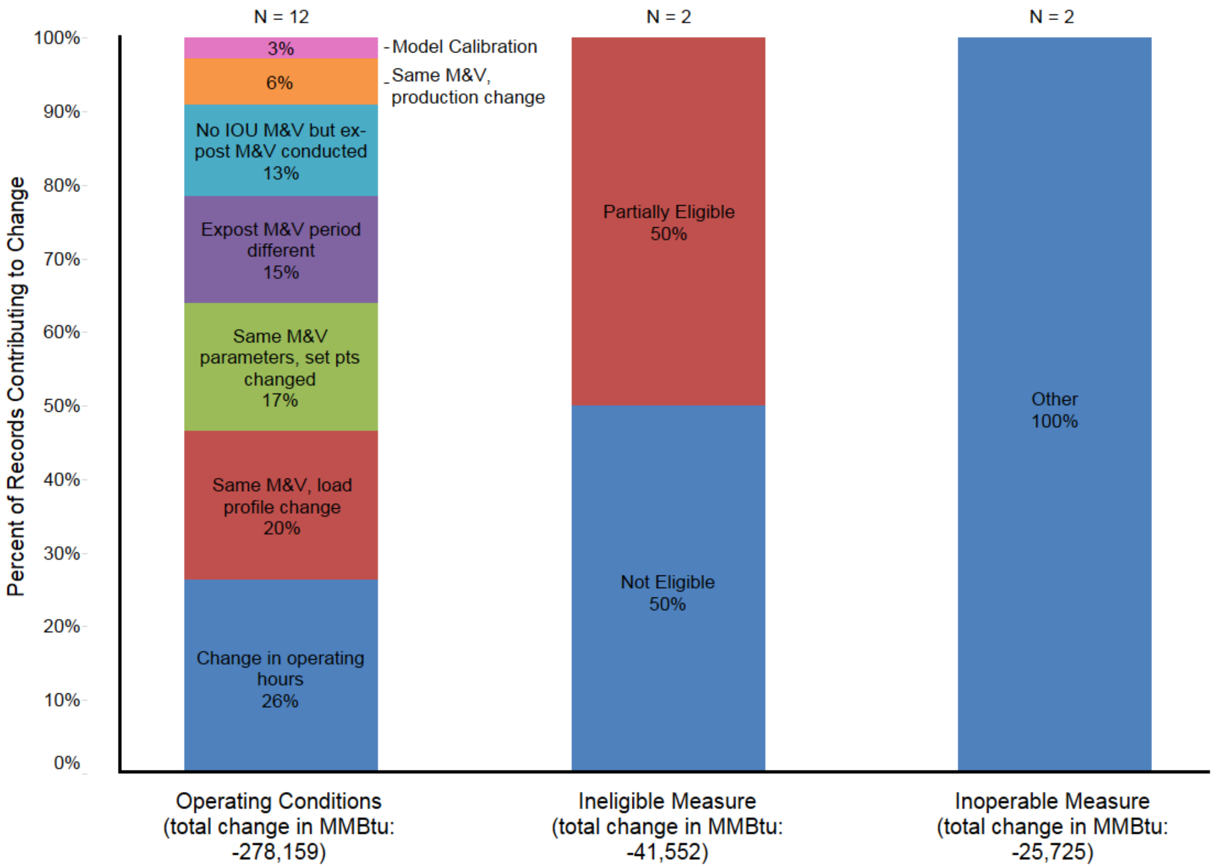


For SDG&E the top three factors that resulted in a downward adjustment of ex-ante MMBtu impacts were operating conditions, calculation methods, and other (Figure 3-9). For 20 records with downward effects due to operating condition, changes during ex-post analysis resulted in an 11 percent reduction to ex-ante MMBtu savings estimates of 333,441. Changes to the ex-post M&V period was the most dominant sub-category, affecting 28 percent of records, followed by set point changes, operating hour changes, load profile changes, use of ex-post M&V, production changes, and model calibration-related changes.

Calculation methods were the second largest reason for downward adjustment to SDG&E's ex-ante MMBtu savings, affecting fourteen records. Sub-categorical explanations were led by changes to inputs and assumptions, followed by calculation method differences and improper PA production normalization.

Two SDG&E records listed as other caused the third largest reason for reduction to SDG&E's ex-ante MMBtu savings, representing a five percent downward adjustment. Both records were associated with a single project where the PA did not adequately document, substantiate and demonstrate a basis for measure savings, including a lack of ex-ante savings calculations.

Figure 3-10: 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 savings were operating conditions, ineligible measures and inoperable measures (Figure 3-10). Changes in operating conditions affected twelve records and resulted in the highest reduction of the ex-ante MMBtu savings estimates. This represents a 37 percent overall reduction to the ex-ante savings of 748,939 MMBtu. Change in measure operating hours was the most frequently observed sub-category within the operating conditions discrepancy group, followed by changes in load profiles, set point changes, ex-post use of a different M&V period, ex-post use of M&V, production changes and model calibration.

There were two records associated with ineligible measures, one of which was determined to be partially eligible. These records resulted in a 5 percent downward adjustment to ex-ante MMBtu savings estimates. For the other ineligible measure the PA had used a regressive baseline that led to ex-post measure rejection. For the partially eligible record one of several measures included in that record was found to be a code requirement and therefore ineligible.

Inoperable measures is the third most important factor reducing SCG's ex-ante MMBtu savings estimates. Two records from a single project were affected by this discrepancy factor, which led to a 100 percent reduction to the associated ex-ante savings. The measures were found to be operable, but not yet fully commissioned or operating as intended/designed, and therefore ineligible. Claiming savings under such circumstances is not appropriate.

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. Those suggestions and considerations are summarized below for each of the main discrepancy factors noted above in Section 3.4 – operating conditions, calculation methods, inappropriate baseline and ineligible measures. The resulting suggestions were examined in all cases where a given discrepancy factor led to a reduction in ex-ante savings of more than 10 percent. Project IDs[61] where relevant suggestions apply are listed in parentheses.

3.5.1 Operating Conditions

Consistent with 2010-2012 and 2013 custom evaluation results, changes in operating conditions represent the single greatest cause for evaluation-based reduction to ex-ante saving estimates. The suggestions reported include some additional steps the PAs and implementers can take to improve ex-ante savings estimates, given that the causes for savings gaps were identified and documented in this evaluation.

M&V Improvement Opportunities

- Ensure that the project equipment are properly identified and that ex-ante savings estimates are based on the appropriate equipment (H40004).
- True-up savings based upon post-installation data, including the use of observed model inputs and conditions. It is noted that within-program M&V may not always be needed for projects but is appropriate, for example, where uncertainty in parameters is a concern (E40021, E40351, E40508, E40511, E40961, E41520, F40002, F40008, F40010, F40011, F40536, F41517, F41531, G40001, G40002, G40039, G40501, G40502, G40505, G40508, G40510, H40006, H40502, H40504, H40506, H40519, H40592, H41522).
  - 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.

[61] Project identifiers include a letter designation that refers to each PA. E for PG&E, F for SCE, G for SCG and H for SDG&E.

   - For example, for EnergyPro simulation models, thorough post-installation M&V should be conducted, including consistency checks between the baseline and proposed equipment and all modeled set points.
- The PAs should be more conservative when estimating savings, given that operating parameters can change and that pre-installation-based parameters and forecasted operations are not always indicative of post-installation conditions. Assumptions and performance of systems should be verified (E40007, E40252, E40352, E40516, E40524, F40019, H40523, H41502).
- PAs should use trend data to generate performance curves instead of using default curves (F40526).
- Ensure that the data collected during the M&V period is representative of typical production or equipment operation. 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 (E40510, E40555, F40242, F40504, H40003, H40013, H40034).

Production Changes

- Before submitting the final savings, the PAs should normalize for production fluctuations (and other non-routine factors and parameters, like weather, pressure settings, etc.) between pre- and post-installation periods, or ensure that pre- and post-installation production levels are comparable (E40252, F40002, F40504, F40547).

Changes in Operating Hours

- Operating hours should reflect observed conditions following equipment installation; verification should be feasible at the time of post-installation inspection or M&V. Also, ex-ante savings estimates should be trued-up based on observed operating conditions following installation. Conduct due diligence to ascertain the annual operating profile of equipment is based on representative data, especially for variable loads, including seasonal variation in production (F41041, F40185, G40501, H40502, H40004).
- For chain accounts installing a measure in multiple locations, operating hours should reflect each unique facility versus the use of a mean for the population or sub-population of facilities (G40578).
- In order to claim kW savings, verify that equipment operates during the coincident peak period using the DEER definition of peak (F40635).

Measure-Specific Issues

- For pump efficiency improvement projects, use actual pump efficiency tests, not estimates. For projects that involve the installation of a VFD it is important to verify that the associated pump will operate at reduced speeds that lead to savings and that controls are in place and operating properly (F40150).
- For applicable oil well field pumping projects, the calculations should be adjusted to represent the post-install discharge pressure, pump depths and fluid levels. Use of a longer period for post-retrofit M&V may be warranted. A combination of SCADA and DOGGR data can be used to best represent post-installation operating conditions. Also non-static efficiencies for pumps and motors should be used in response to changes in observed loads (F40025, F40557).
- For VFDs it is suggested that the PAs take spot power readings at full speed to verify fan load factor instead of using assumed values. Also, the control strategy for the VFD should be documented (E40003).

3.5.2 Calculation Methods

Ex-Post Calculation Method Different from PA

- PAs should use acceptable industry standard tools that are capable of addressing all impact estimation details, such as the capturing of interactive effects across measures in whole-building simulations (E41503).
- PAs should use transparent savings calculation tools when appropriate, such as a simple spreadsheet models. Use of proprietary and "black box" tools prevents the evaluation from building upon PA methods and explaining discrepancy factors that lead to differences between ex-ante and ex-post estimates of savings (G40018, H40022).
- The PAs should ensure that projects have an identifiable and documented case for energy efficiency claims. This should include the identification of energy improvements being made to equipment and an explanation for how the project leads to savings, especially for projects that are unique or for projects that involve complex systems or processes (E40603, H40513).
- For agricultural pumping upgrade projects, the OPE methodology should not be used when the pump lift (head) changes by more than 10% (E40786).
- Peak demand savings estimates should be based upon the DEER-defined peak demand reduction period (E40786).
- Where available PAs should use metering, EMS or SCADA data to confirm or derive model inputs, such as operating conditions and condensate return temperatures. Where available these data should also be used to calibrate models. In the absence of defensible

data, use conservative estimates for model inputs; for example, when calculating theoretical baseline operation in the absence of measured data (F40501, F40893, G40526, H40001, H40508).

- PAs should ensure that calibration data are representative of typical operations, and use of back-casting where needed (E40503).
- Provide a fully unlocked savings model (G40512).
- Equipment loads and profiles are critical inputs needed for savings estimation, and should be a key data input collected by the PAs (G40520).
- Perform a sanity check using simple rules-of-thumb in order to be confident in calculated results. For example, verify motor loads that fall outside of a normal range (H40003).
- For NRNC whole-building projects the PAs should use non-compliance mode to estimate savings (H40083).
- Follow the guidance document posted by the CPUC on estimating savings when non-IOU supplied energy sources are used. For example, performing an hourly net grid impact analysis if onsite generation is significant (H40529).

Errors Found in PA Calculation Model

- Ensure that savings developed using calculation tools are reflected appropriately in the tracking system claim (E40002).
- For certain projects it is appropriate to verify savings by comparing baseline and post-installation system-level specific power, such as kW/CFM (E40004).
- Evaluators should be able to replicate PA savings estimates using the same model and inputs (E40005).
- PAs should select tools that are appropriate for a given measure and ensure that savings calculations are conducted correctly, including validation that the tool works properly and that inputs are appropriate (E40003, E40620, H40523).
- Where appropriate PAs should double-check that IR approved baselines are used in savings estimation models (F40624).

Same Calculation Methods, Inputs and Assumptions Changed

- PAs should confirm and use actual equipment specifications and model inputs. This should include the true-up of post-implementation models with as-built equipment specifications, as well as temperature settings and equipment operating schedules. If available, air balance reports and Cx tests can be referenced for major building equipment (E40016, E40507, F40451, H40001, H40565).
- The PAs should account for all equipment and system efficiencies when developing savings estimates, and review formulas for completeness, for example, accounting for VFD efficiency (F40551).

- For projects entailing weather sensitive measures, the PAs should verify that the correct weather files are used in the analysis (F40624).

Incorrect Methods Used for Saving Normalization

- PA use of regression techniques should ensure that modeling and results are robust (E40510).

3.5.3 Inappropriate Baseline

Wrong ISP/code Corrected with Right ISP/code

- The PAs should carefully investigate and document the age, condition and functionality of existing equipment and operations, and use these to establish proper baselines. A baseline must be established that can meet the service requirements under post-installation conditions (e.g., pumping requirements). This is especially true for capacity expansion projects, where 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 sized adequately and otherwise able to meet loads and operating condition requirements, in addition to having efficiency levels that are above code or ISP (E41163).
- For all industrial steam boiler projects that include efficiency upgrades, use 82% minimum baseline combustion efficiency until relevant ISP research is performed (E40516).
- For new steam generator applications in oil production, consultation with industry experts has identified optimized split-flow generator configurations, with fan and pump VSDs, to be ISP. The PAs should carefully consider ISP for all projects and applications and update market assessments frequently. In the absence of rigorous ISP evidence, being conservative is warranted with respect to baseline determination (E40555, F40536).
- For NRNC whole-building projects the baseline conditions that are automatically generated by EnergyPro are not always consistent with the ACM manual and SBD modeling procedures. The PAs should examine all parameters and make corrections to the baseline model where needed (E40351).
- When choosing baseline equipment for certain ROB/NR projects, the PAs should set efficiency levels for equipment of the same type and capacity as the newly installed equipment. For example, in HVAC applications where Title 24 minimum efficiency levels are used to establish baseline, and NR replacement of burners (F40010, G40628, G40632, G40639).
- 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. A careful examination is warranted to establish design options that

are available to the customer, and to establish that the program-supported equipment solution is a legitimate high efficiency action (F40001).

- During application review the PAs should carefully consider all relevant code requirements and update ISP determinations for relevant measures. For example, insulation is a safety requirement under Cal-OSHA for specific tank applications, ASHRAE 90.1may offer the relevant code requirement, and federal codes mandate ASHRAE 90.1-for HVAC equipment (F40624, G40004, G40012, G40513, H40505, H40505B).

ER Overturned to ROB/NR

- For ER claims the condition and functionality of the existing equipment should be documented, along with the RUL. Preponderance of evidence that the program induced the early retirement should be used to establish early retirement claims (F40517, H40034).

Other

- Ensure that baseline equipment operational parameters are properly documented and modeled (F40526, H41505).
- Applications submitted by the PAs should document both the project type and baseline for all claims (F40242).
- When baseline conditions are defined by the pre-existing system the PAs should utilize measured data to define those conditions, and thoroughly document the pre-existing conditions for the purposes of establishing baseline. For example, assuming a particular flow rate and/or other conditions that define capacity for the baseline system, or using an assumed schedule and settings for HVAC equipment (E40003, E40011, F40008).
- For ROB claims it is improper to use the pre-existing equipment as the baseline, except for the purposes of setting baseline efficiency levels when the pre-existing equipment is more efficient than code/ISP (regressive baseline rule). By definition ROB project baselines involve the removal of the existing equipment (F40052, H40003).

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, and wastewater solutions that exceed the efficiency levels of prior practices. Otherwise the installation constitutes a like-for-like replacement, which is not considered to be an energy efficiency action (E40587, E40524).
- The PAs should document pre-existing equipment and conditions for all retrofit projects to ensure that the new equipment efficiency level is high enough to qualify for incentives. For replacement measure retrofits, the new condition must be more efficient than both the baseline condition and the pre-existing condition. For example, replacing insulation with

an equivalent amount of insulation is not program qualifying because retrofitting with a measure that is of equivalent efficiency is not an allowable energy efficiency action (E40520, E40587).

3.5.4 Ineligible Measure

- To properly screen projects for eligibility the PAs should document project acceptance dates and carefully review ISP determinations and effective dates (E40030, E40536).
- PAs should confirm that projects meet program eligibility performance thresholds, for example, the Savings by Design time dependent valuation (TDV) savings margin (E41555).
- Models submitted by the PAs for determining energy savings estimates should reflect as-built conditions. A prototype model that represents a dramatically different design should not be eligible for incentives (F40452).
- The PAs should ensure that projects have an identifiable and documented case for energy efficiency claims. This should include the identification of energy improvements being made to equipment and an explanation for how the project leads to savings, and should not include operational changes that are ineligible for incentives (F40502).
- Whenever determining energy savings at a facility that has onsite generation, actual PA grid impacts must be taken into account, and the amount of generation/cogeneration versus PA grid energy consumption must be correctly attributed in ex-ante savings estimates, including hourly analysis. Only partial savings were allowed where hourly impacts exceeded PA imports (H40002, H40009, H40501, H40509).
- The PAs should document that the installed measures exceed code/ISP baseline performance levels and do not entail like-for-like replacements, or regressive baselines. This includes the requirement that measure installations be screened to be sure they exceed code requirements at the time of construction or subsequent upgrade (F40017, F40010B, F40525, F40629, G40504, G40570, H40011, H40048, H40506).

4

NTG Results

The methodology used to develop individual, site-specific net-to-gross (NTG) estimates is summarized in the IALC Research Plan.[62] Weighted NTG results are presented in this chapter 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.[63]

4.1 Number of Completed Surveys

One hundred and ninety-six NTG surveys were completed in total. The original sample design consisted of 150 sample points that overlap with the Gross impact M&V sample design plus an additional 50 NTG-only sample points, which were evenly distributed across the size strata. However, given customer willingness to participate and other factors, the final gross and net samples did not fully align. In total, 116 of the completed NTG sample points overlapped with the 150 evaluated gross M&V points. 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. Each utility accounted for roughly one-quarter of the completed surveys, with SCE having the largest number of completes (57) and SCG having the least (42). The surveys completed represent 36 percent of ex-ante MMBtu claims.

Table 4-1: Completed Surveys by Program Administrator

Program Administrator	Completed NTG Points (n)		Percent of Ex-Ante MMBtu Claims
	Main	Backup	NTG Sample
PG&E	42	10	37%
SCE	46	11	29%
SDG&E	34	11	43%
SCG	35	7	38%
All PAs	**196**		**36%**

62 http://www.energydataweb.com/cpuc/deliverableView.aspx?did=1133&uid=0&tid=0&cid=

63 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

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 savings estimates (savings) associated with the application and the proportion of the total sampling domain savings represented by each sampling stratum. Since the sample of electric and gas projects was developed based on one common metric, source Btu, NTGR results are weighted 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 utility-level only.

Note that the final NTGR values in Table 4-3, Table 4-4, Table 4-5 and Table 4-6 below are based on the removal of 4 surveyed projects, leaving a total of 196 sample points. This was due to either an ineligible measure (3 projects removed) or inconsistent answers between the EAR and post-installation survey responses (1 project removed). In addition there were another six projects for which NTG data collection efforts were canceled based on early gross impact findings. These six projects were replaced with backup points. In general these ten projects (Table 4-2) were excluded from the NTG analyses in order to avoid double-counting of downward adjustments to project savings across both the M&V and NTG efforts (for the same reasons).

Table 4-2: Surveyed Projects Removed from NTG Analysis

ItronID	Reason for Removal
E40536	The measure was found to be ineligible, therefore, there is no legitimate project to evaluate.
E41555	The measure was found to be ineligible, therefore, there is no legitimate project to evaluate.
F40525	The measure was found to be ineligible, therefore, there is no legitimate project to evaluate.
F40524	This project was removed from the NTG sample because of inconsistencies in survey responses that could not be resolved.
F40001	NTG data collection efforts were canceled because of early gross impact findings
F40017	NTG data collection efforts were canceled because of early gross impact findings
G40502	NTG data collection efforts were canceled because of early gross impact findings
F40502	NTG data collection efforts were canceled because of early gross impact findings
G40504	NTG data collection efforts were canceled because of early gross impact findings
E40533	NTG data collection efforts were canceled because of early gross impact findings

4.2.1 PG&E Combined Electric and Gas

Table 4-3 below reports NTGRs results for PG&E. The resulting NTGR of 0.51 for 2014 is down slightly from the 2013 NTGR value of 0.55.

Table 4-3: Weighted Net-to-Gross Ratios for PG&E – Combined Electric and Gas

Sampling Strata	MMBtu NTGR	N Sample Frame	n NTGR Sample	Percent of MMBtu Sampled
1	0.60	2	2	100%
2	0.56	22	16	71%
3	0.60	34	12	37%
4	0.52	135	10	7%
5	0.31	1,051	12	1%
All - Weighted NTGR	**0.51**	**1,244**	**52**	**37%**
90 Percent CI	0.47 to 0.54			
Relative Precision	0.06			
n NTGR Completes	52			
N Pop Sampling Units	1,244			
ER	0.29			

Results by Stratum: Results varied across sample size strata, with stratum 5 exhibiting the lowest NTGR (0.31), while stratum 1 through stratum 4 NTGRs were clustered in the 0.52 to 0.60 range.

- Stratum 1, consisting of the largest projects, had the highest NTGR along with stratum 3. The weighted mean NTGR is the result of a new construction hospital building and a food manufacturer with NTGRs of 0.47 and 0.67
- Stratum 2 projects had a wide range of results ranging from 0.05 to 0.90. This size stratum consisted of many oil refinery and oil well projects but also included a few new construction, manufacturing, laundry, casino and university projects. Results among the oil company projects were mixed, and ranged from lows of 0.43 (2 projects) to 0.90 (3 projects).
- Stratum 3 projects include a mix of oil, new construction, food/agriculture, data center and institutional sector projects. NTGRs ranged from 0.23 to 1.00 with a majority of project-level results being above 0.50. Only 4 of the 12 projects evaluated had NTGRs below this level.
- Strata 4 and 5 consisted of a wide mix of smaller projects. Noteworthy results included much lower NTGRs among the oil, hospital new construction, manufacturing, technology, and agriculture projects in particular. These results yielded a weighted average NTGR of

0.31 for stratum 5 and 0.52 in stratum 4. Both results were low relative to results in the other strata.

4.2.2 SCE Electric

Table 4-4 presents SCE NTGR results. The resulting weighted average 2014 NTGR is 0.46, and has decreased from 0.57 based on 2013 results.

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

Sampling Strata	MMBtu NTGR	N Sample Frame	n NTGR Sample	Percent of MMBtu Sampled
1	0.42	8	6	76%
2	0.47	37	14	34%
3	0.49	62	14	24%
4	0.48	131	12	8%
5	0.41	923	11	1%
All - Weighted NTGR	**0.46**	**1.161**	**57**	**29%**

90 Percent CI	0.42 to 0.49
Relative Precision	0.07
n NTGR Completes	57
N Pop Sampling Units	1,161
ER	0.35

Results by Stratum. Average NTGR results across all strata range from 0.41 to 0.49. The mid-sized projects in strata 2, 3 and 4 have somewhat higher NTGRs than the strata 1 and 5 projects. Specific drivers are discussed below:

- Stratum 1 project NTGRs ranged from 0.20 to 0.70. Four of the six projects in this size stratum had NTGRs of 0.50 and higher, reflecting medium program influence. However, the two remaining projects had extremely low NTGRs of 0.20 and 0.32 and accounted for nearly 40% of the stratum savings.
- Stratum 2 project NTGRs varied widely. Half of the projects (7 of 14) had NTGRs of 0.50 and above signifying medium program influence. The remaining half of the projects (7 of 14) had NTGRs of 0.44 and below, including 3 projects with very low NTGRs of 0.00, 0.29 and 0.31. These 3 projects accounted for 18% of stratum 2 ex-ante MMBtu savings in the sample.
- Stratum 3 consisted of a mix of legacy smart well projects, as well as projects installed by national and grocery chain stores, telecommunications, aerospace and small commercial businesses. Nine of the 14 projects in this stratum exhibited medium to high levels of

program influence, with NTGRs of 0.50 and above (including one project with a perfect NTGR of 1.0). The remaining 5 projects were primarily national and grocery chain stores and all had NTGRs of 0.34 and below. These 5 projects represented 41% of ex-ante savings and pulled the overall result down significantly.

- Stratum 4 included several strong performing projects with high program influence (NTGRs of 0.65 and higher), and other poor performing projects, including several that were undertaken by municipal water agencies. Those in the latter category generally had low NTGRs (0.37 and below) and municipal water projects were flagged in previous Custom program evaluations as being prone to high free ridership.
- The majority of projects in Stratum 5 (8 of 11 projects) had NTGRs of 0.40 and below, signifying low program influence. Among these were six projects with NTGRs of 0.33 and below. Collectively, these projects' NTGRs explain the low average NTGR of 0.41 for this size stratum.

4.2.3 SDG&E Combined Electric and Gas

Table 4-5 presents 2014 NTGR results for SDG&E. The average 2014 NTGR for SDG&E's electric and gas projects is 0.51. This represents a decline from the SDG&E 2013 NTGR result of 0.59.

Table 4-5: Weighted Net-to-Gross Ratios for SDG&E – Combined Electric and Gas

<table>
<tr><th>Sampling Strata</th><th>MMBtu NTGR</th><th>N Sample Frame</th><th>n NTGR Sample</th><th>Percent of MMBtu Sampled</th></tr>
<tr><td>1</td><td rowspan="2">0.56</td><td>2</td><td>1</td><td>43%</td></tr>
<tr><td>2</td><td>9</td><td>6</td><td>59%</td></tr>
<tr><td>3</td><td>0.50</td><td>19</td><td>10</td><td>53%</td></tr>
<tr><td>4</td><td>0.44</td><td>34</td><td>13</td><td>37%</td></tr>
<tr><td>5</td><td>0.51</td><td>139</td><td>15</td><td>11%</td></tr>
<tr><td>All - Weighted NTGR</td><td>0.51</td><td>203</td><td>45</td><td>43%</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>90 Percent CI</td><td>0.47 to 0.56</td></tr>
<tr><td>Relative Precision</td><td>0.09</td></tr>
<tr><td>n NTGR Completes</td><td>45</td></tr>
<tr><td>N Pop Sampling Units</td><td>203</td></tr>
<tr><td>ER</td><td>0.42</td></tr>
</table>

Results by Stratum. Results varied somewhat across sample size stratum ranging from a low of 0.44 (stratum 4) to a high of 0.56 (strata 1 and 2).

- NTGRs of projects in strata 1 and 2 range from 0.30 to 0.85. The majority of projects in these two size strata (4 of 7 projects) had NTGRs of 0.52 and above. The strongest performing projects include two university-based retrocommissioning projects, with NTGRs of 0.85. These were offset somewhat by the remaining 3 of 7 projects with NTGRs below 0.43 and which accounted for 49% of ex-ante MMBtu savings among projects that were sampled in strata 1 and 2.
- Stratum 3 projects include school, hotel, technology, manufacturing and hospital projects. NTGRs ranged from 0.18 to 0.85, with half of the sample yielding a result less than or equal to 0.50.
- NTGRs for smaller projects, particularly those in Strata 4 and 5, ranged in values from a low of 0.13 to a high of 0.74. The average NTGRs for each of these strata represents the mid-point of the range of values of contributing projects.

4.2.4 SCG Gas

For SCG gas projects, the weighted NTGR result is 0.62, as shown in Table 4-6 below. This result is down somewhat from the 2013 estimate of 0.66.

Table 4-6: Weighted Net-to-Gross Ratios for SCG – Gas

Sampling Strata	MMBtu NTGR	N Sample Frame	n NTGR Sample	Percent of MMBtu Sampled
1	0.66	3	1	34%
2		9	3	28%
3	0.62	8	7	89%
4	0.52	25	14	59%
5	0.52	191	17	10%
All - Weighted NTGR	**0.62**	**236**	**42**	**38%**

90 Percent CI	0.60 to 0.65
Relative Precision	0.04
n NTGR Completes	42
N Pop Sampling Units	236
ER	0.18

Results by Stratum. NTGR results by size stratum ranged from a low of 0.52 for strata 4 and 5 to a high of 0.66 for strata 1 and 2.

- NTGR results in the three largest size strata (strata 1, 2 and 3) were dominated by manufacturing and food/agriculture projects exhibiting medium high to high levels of program influence. Results in each of these strata are based upon a relatively small sample size (Stratum 1: one project, stratum 2: three projects, stratum 3: seven projects). All but one of these projects had a NTGR of over 0.50.
- Among smaller sized projects in Strata 4 and 5, there was wider variation in the results of contributing projects. NTGRs ranged from 0.25 to 0.78 in Stratum 4 and from 0.14 to 0.67 in Stratum 5. NTGRs in the low-end generally explain the relatively low average NTGR results obtained in these strata. Stratum 5 consisted of many fast food sector projects but also included manufacturing, oil extraction, and institutional projects. Results for the fast food sector projects were clustered, with project-level NTGRs ranging from 0.55 to 0.57.

4.2.5 Comparison of 2014, 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 with 2014 and 2013, 2010-12 results were weighted by fuel-based MMBtu and all three sets of results are presented in Table 4-7 and Figure 4-1. These exhibits show that 2014 NTGR results are statistically significantly lower than 2013 results at the 90% confidence level, but similar to 2010-12 results.

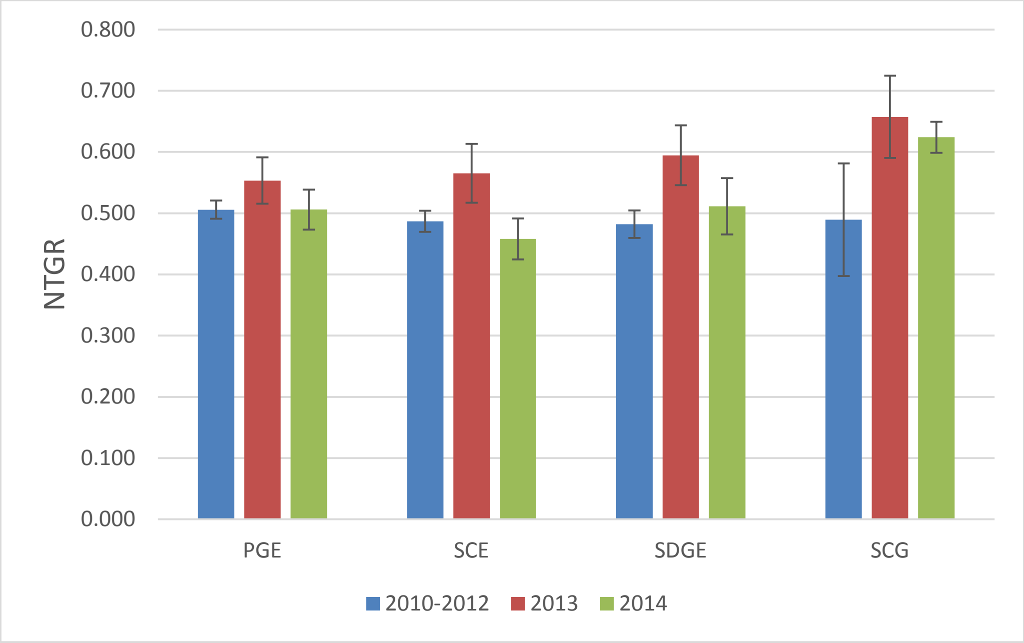
- 2014 NTGRs by PA range from 0.46 to 0.62.
- **PG&E:** 2014 weighted NTGRs results for PG&E (0.51) declined by 9 percent compared to 2013 evaluation results (0.55) but are the same as 2010-12 results (0.51).
- **SCE:** 2014 NTGR results for SCE (0.46) declined by 19 percent compared to 2013 results (0.57) but are comparable, though somewhat lower than, 2010-12 results (0.49).
- **SDG&E:** 2014 NTGR results for SDG&E (0.51) declined by 14 percent compared to 2013 results (0.59) but have improved by 6 percent compared to 2010-12 results (0.48).
- **SCG:** For SCG the 2014 weighted NTGR across all projects is 0.62. This represents a 5 percent decline compared with 2013 average NTGR results of 0.66. However, 2014 results have improved by 28 percent relative to 2010-12 results (0.49).

Table 4-7: Comparison of 2010-12, 2013, and 2014 Weighted MMBtu* NTGR Results

Energy Metric	2010-2012 Mean NTGR	2010-2012 90% Confidence Interval	2013 Mean NTGR	2013 90% Confidence Interval	2014 Mean NTGR	2014 90% Confidence Interval
PG&E						
MMBtu*	0.51	0.49 to 0.52	0.55	0.52 to 0.59	0.51	0.47 to 0.54
SCE						
MMBtu*	0.49	0.47 to 0.50	0.57	0.52 to 0.61	0.46	0.42 to 0.49
SDGE						
MMBtu*	0.48	0.46 to 0.50	0.59	0.55 to 0.64	0.51	0.47 to 0.56
SCG						
MMBtu*	0.49	0.40 to 0.58	0.66	0.59 to 0.73	0.62	0.60 to 0.65

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

Figure 4-1: Comparison of 2010-12, 2013 and 2014 Weighted MMBtu* NTGR Results



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

4.3 NTG Sensitivity Analysis

NTG ratios were calculated for each of the sample points based equal weighting for each of the three scores. A sensitivity analysis of the resulting NTGRs was 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 4-8. Note that unlike the other four weighting schemes, only scheme 6 relies on just one of the 3 scores under certain conditions, and only scheme 3 takes a mean of just two of scores.

Table 4-8: Results of NTG Sensitivity Analysis – All PAs and All Sample Points

NTGR Weighting Scheme	NTGR Result*
1. 33.3% weights to scores 1, 2 and 3 (current approach)	0.52
2. 50% weight to score 1, 25% to scores 2 and 3	0.52
3. Remove score 1, 50% weight to scores 2 and 3	0.51
4. 50% weight to score 2, 25% to scores 1 and 3	0.51
5. 50% weight to score 3, 25% to scores 1 and 2	0.52
6. Use only score 3 if no-program likelihood is 10	0.48

* Based on simple averaging.

These results indicate that the resulting NTGR results are not very sensitive to the weighting scheme used. Only the extreme case (6), when score 3 was used exclusively if the no-program likelihood of installation was scored a 10 out of 10, showed variation and this was relatively minor.

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

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. 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 4-9 presents the results of this analysis across all projects where respondents offered the strong importance scores of 8, 9 or 10 for a given project driver (on a scale of 0 to 10). The NTGR information in the top four rows of the table provides a breakdown of each utility's evaluated NTGR by quartile. High percentages of projects in the Medium-Low and Low quartiles (the two right-most columns) are of greatest concern with respect to free ridership. The Key NTG Project Drivers information in the remainder of the table provides insight into the factors that drive free ridership. The percentages indicate the frequency with which respondents assigned a given project driver a strong importance score (8, 9, or 10) within each NTG quartile. By examining these percentages, some correlations between important project drivers and high or low NTG values can be observed. For each driver, the directional influence on the NTGR by each of the factors is also indicated. Strong importance scores for “positive” drivers are hypothesized to lead to high NTG values whereas strong scores for “negative” drivers are hypothesized to lead to lower NTG values.

Table 4-9: Key Factors Affecting NTGRs for all PAs

	Influence on NTGR	Distribution of NTGRs			
		High - 0.76 to 1.00	Medium High- 0.51 to 0.75	Medium Low- 0.26 to 0.50	Low - 0.00 to 0.25
IOU					
PG&E*		15%	35%	33%	17%
SCE*		4%	33%	47%	16%
SCG*		5%	69%	19%	7%
SDG&E*		9%	42%	40%	9%
Key NTG Project Drivers					
Project Maturity					
Decision to install was made after learning about the availability of rebates through the program	Positive	88%	87%	38%	16%
Corporate Policy/Practice					
Measure is part of corporate standard practice	Negative	13%	47%	43%	56%
Company has corporate policy in place that influenced the project	Negative	69%	52%	68%	64%
Energy Efficiency A Secondary, not Primary, Benefit					
Measure was added to control or work directly with existing equipment	Negative	31%	41%	22%	20%
Energy efficiency was a motivation	Positive	94%	67%	54%	52%
Measure improves workplace safety	Negative	13%	13%	7%	12%
Market Segment					
Measure is installed by a market segment that is ahead of curve on energy efficiency	Negative	25%	0%	0%	4%
Measure is installed by national chain/big box firm	Negative	19%	14%	13%	4%
Measure is installed by single location business	Positive	13%	22%	10%	4%
Project Cost vs. Rebate					
Rebate is high % of first year project cost (> 25%)	Positive	81%	68%	59%	72%
Project Context					
Measure is part of an expansion/remodeling	Negative	31%	20%	39%	20%
Measure installed to replace old or failing equipment	Negative	19%	29%	36%	52%

* Percentages in the columns at right show the frequency of NTGRs in each range (high, medium high, medium low, and low) within each PA.

The following are general themes and observations that emerged based on this analysis:

- Certain factors are correlated with a **lower level of program influence**, and include those listed below.
  - Corporate standard practice (i.e., the measure is installed elsewhere at other locations by the customer, including those without rebates)
  - Measures installed to replace old or failing equipment without the explicit demonstration of program influence on selection
- Other factors that can contribute to **higher program influence** at the project level include:
  - Timing of decision follows discussions with program staff
  - Energy efficiency is a key motivation

The hypothesized directional impact of the various factors are based not only on what logic would dictate, but also on patterns observed in recent evaluations. However, the data above are not consistent with these directional impacts in all cases. Because of the limited sample size, the findings could not be stratified in sufficient detail to tease out the effects of other project drivers, if present.

As noted in Chapter 6, Findings and Recommendations, additional levels of program influence can be achieved through the awareness of and by taking action in response to these and other important factors.

5

Project Practices Assessments

5.1 Introduction

Project Practices Assessments (PPAs) are structured site-specific reviews of Program Administrator (PA)[64] 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 Project Practices Assessment was first conducted in the 2013 IALC evaluation. 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.[65] The PPA results are a companion to the Chapter 3 gross impact results. For example, PPA findings help to explain discrepancy factor results that lead to upward and downward adjustments to ex-ante savings estimates, based on cross-project differences in conclusions by the evaluators. PPA findings also identify critical weaknesses in documentation and reporting.

---

64 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.

65 By developing results that are presented in a quantitative format, it will be feasible to use 2013 results as a baseline and measure PA trends that emerge relative to that baseline.

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-ante reviews).[66] Pre-2013 results serve as an initial baseline against which to measure 2013+ differences.[67]

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
- Project type selection
- Baseline selection
- Project EUL assessment
- 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 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.

The PPA form and procedure was designed to document both the PA and evaluator conclusions and to ensure that results could be analyzed objectively to assess conformance with policy guidelines, best practices and program rules.

---

66 http://docs.cpuc.ca.gov/published/FINAL_DECISION/139858.htm
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.

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

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

- Quality and comprehensiveness of documentation
- Accuracy and appropriateness of ex-ante inputs, assumptions, results and conclusions

A 1 to 5 rating scale is used to examine criteria on each project and measure-specific PPA form, 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. It should be noted that these are quantitative scores meant to capture a range of qualitative information and are, therefore, somewhat subjective. The evaluation team made every effort to ensure consistency across PPA scoring, including a PPA consistency check by a single engineer for all evaluated measures. Appendix E presents full scoring guidelines used by the evaluation team.

5.3 Project Practices Assessment Results

This report presents unweighted PPA results by PA. The results are further aggregated for measure-level records with agreement dates that are pre-2013 versus 2013+. It is noteworthy that resulting sample sizes for the pre-2013 period are small for three of the PAs: PG&E, SCG, and SDG&E, at just 15 measures, 12 measures and 14 measures, respectively. These sample sizes are too small to yield conclusive results, and results throughout this section should be reviewed with that key consideration in mind. In total, 188 individual records from 150 projects were evaluated across all PAs.[68] 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 PA-specific results and findings from the PPA analysis can be found in Appendix F.

5.3.1 Project Eligibility Assessments

Table 5-1 presents the PPA findings regarding project eligibility by PA and time period (pre-2013 and 2013+). Each record in the tracking data was classified as either eligible, ineligible, or partially ineligible. The partially ineligible designation arises in the case where a given record, which typically comprises a single measure, is actually comprised of multiple measures that have one or more ineligible components. For each time period and PA, Table 5-1 displays the number of ineligible and partially ineligible measures. The table also presents the ex-post conclusions for

[68] Note that there were actually 192 records associated with the 150 gross M&V points. Four of these records (two from SCG and two from SDG&E) were removed from the PPA analysis because they are "incentive only" measures that are not subject to engineering review.

why measures were determined to be ineligible. While a variety of reasons for ineligibility were cited, the most common reason for ineligibility is that measures do not exceed the code or ISP baseline (18 of 38 reasons cited). The remainder of ineligible or partially ineligible projects were due to CPUC guidance and decisions, previous EAR guidance, non-PA fuels and ancillary impacts, program rules, and stipulations in the EE policy manual. These same issues were identified in Chapter 3 as the causes for eligibility issues that led to substantial downward adjustments to ex-ante gross MMBtu savings estimates. For example, greater effort is needed on the part of the PAs to screen measures to ensure that they exceed code/ISP requirements. Also, greater levels of communication are needed with PA staff and contractors involved in implementing custom offerings, to ensure conformance with CPUC eligibility guidance. This includes improvements that should be made to PA program requirements, manuals, training, and quality control procedures.

Table 5-1: Summary of INELIGIBLE MEASURES by Customer Agreement Date, and Ex-Post M&V Conclusions Why Measures are INELIGIBLE

Parameter Examined	PA Eligibility Treatment							
	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	15	32	12	14	33	24	20	38
Number of INELIGIBLE Measures	0	4	0	1	2	2	1	2
Number of Partially INELIGIBLE Measures	2	0	0	7	4	2	1	2
Percent of Measures Found to be INELIGIBLE	0%	13%	0%	7%	6%	8%	5%	5%
Percent of Measures Found to be Partially INELIGIBLE	13%	0%	0%	50%	12%	8%	5%	5%
Evaluation Conclusions Why Measures are INELIGIBLE or Partially INELIGIBLE								
Conclusions								
Program rules	0	1	0	0	3	0	0	0
Normal maintenance	0	0	0	0	0	0	0	0
Operating practice change	0	0	0	0	0	0	0	0
CPUC decisions	0	1	0	0	0	0	0	0
CPUC guidance	0	1	0	1	2	1	1	0
Requirement that measures exceed code / ISP baseline	2	3	0	2	4	3	2	2
Previous EAR guidance	0	0	0	1	1	0	0	0
Previous evaluation findings	0	0	0	0	0	0	0	0
Project boundary condition	0	0	0	0	0	0	0	0
EE Policy Manual	0	1	0	0	0	0	0	0
Multiple PA fuels (includes cogeneration and fuel switching)	0	0	0	0	0	0	0	0
Three prong test	0	0	0	0	0	0	0	0
Non-PA fuels and ancillary impacts (i.e., cogen, refinery gas, WHR, etc.)	0	0	0	4	1	0	0	1

PA-specific findings include the following:

- SCG had few eligibility issues for 2014 projects overall, affecting just 2 out of 34 measures evaluated across both the pre-2013 and 2013+ periods. This aligns with findings from Chapter 3 where it was shown that just one project consisting of two measures, G40504, experienced eligibility issues. In fact, only this project and one other SCG project had zero ex-post MMBtu gross savings out of the entire sample of 30 projects.
- Although SDG&E eligibility treatment in the pre-2013 period was well below expectations (given that eligibility issues were identified in 8 out of 14 measure evaluated), eligibility treatment in the 2013+ period was well above average, affecting just 4 out of 38 measures evaluated. SDG&E also only had two projects that were identified in Chapter 3 as having zero ex-post MMBtu gross savings out of the entire sample of 35 projects.
- In Chapter 3 a total of 22 projects were identified as having zero or negative ex-post MMBtu gross savings, 9 each for PG&E and SCE. About half of those 18 projects and a roughly equal number for each of these PAs had eligibility issues that led to such low ex-post savings. There is considerable room for improvement in eligibility treatment for these two PAs, as well as greater attention to correcting project issues associated with low evaluated savings. As was noted in chapter 3, eligibility was the largest SCE discrepancy factor leading to downward adjustments to ex-ante savings estimates.

5.3.2 Project Type Assessment

Establishing the correct project type (retrofit add on, early replacement/retirement, normal replacement, replace on burnout, capacity expansion, new construction, major renovation or system optimization) is a first order consideration in the project application process. Project type has important implications for baseline selection, the use of incremental and/or full costs, proper application of relevant codes and standards, the applicability of EUL and RUL, and first year and second baseline period savings calculations. In particular, baseline selection and treatment can be impacted by improper project type designation. It was noted in Chapter 3 that inappropriate baseline selection was one of the leading causes for downward adjustments to ex-ante MMBtu gross savings estimates. So it is important to properly document project type from a gross impact perspective.

While perfect agreement between PA and evaluator specified project types reduces the likelihood of evaluated savings deviating 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, a PA project classified as an add-on measure may be reclassified as system optimization without any impact on first year savings because the baseline for both project types is ordinarily the pre-existing system.

However, an ER project evaluated as a ROB project or vice versa can significantly impact the FY and LC GRRs.

Table 5-2 presents the frequency of ex-ante and ex-post agreement (and disagreement) on project type by PA and time period (pre-2013 customer agreement date versus 2013+). For all PAs combined, there was some improvement in project type selection in the 2013+ period relative to the pre-2013 period (pre-2013: 59 percent agreement vs. 2013+: 64 percent agreement).

Table 5-2: Frequency of Ex-ante and Ex-Post Agreement on Project Type by PA and Customer Agreement Date

PA	Customer Agreement Date	Total Project Types Examined	Project Types Matched	Project Types Overturned	Project Types % Matched
PG&E	Pre-2013	15	11	4	73%
	2013 +	33	24	9	73%
SCE	Pre-2013	32	19	13	59%
	2013 +	24	7	17	29%
SCG	Pre-2013	12	6	6	50%
	2013 +	20	15	5	75%
SDG&E	Pre-2013	14	7	7	50%
	2013 +	38	28	10	74%
All PAs	Pre-2013	73	43	30	59%
	2013 +	115	74	41	64%

PA-specific findings include the following:

- PG&E’s accuracy on project type was generally quite good, being correctly identified 73 percent of the time in both the pre-2013 and 2013+ periods. It is notable, as shown below in Section 5.3.3, that PG&Es accuracy on baseline specification was also consistently high, and some level of correlation would be expected between the two parameters.
- SCE’s accuracy on project type selection was poor overall and especially low in the 2013+ period, where agreement with the evaluation conclusion was only achieved 29 percent of the time. As noted below in Section 5.3.3 SCE also achieved low accuracy in baseline specification, as might be expected due to the co-dependency of these two parameters.

- SCG and SDG&E demonstrated considerable improvement in the accuracy of project type designation in the 2013+ period relative to pre-2013, where both PAs improved by approximately 25 percent relative to pre-2013.

Table 5-3 presents results detailing ex-ante versus ex-post project type designations for all PAs. The green shaded cells along the diagonal indicate the number of measures that showed agreement between the PA and ex-post evaluation. Values in the red shaded cells are measures where the project type was reassigned by the evaluator. As stated previously , the most commonly overturned project types were add on, early replacement, and replace-on-burnout measures. In the pre-2013 period, PA assigned add-on measures were overturned 42 percent of the time (to NR, ROB, and SysOp), early replacement measures were overturned 86 percent of the time (to NR, ROB, SysOp, and add-on) and ROB measures were overturned in 50 percent of cases (to add-on, ROB, NR, and SysOp). In the 2013+ period, add-on, ER, and ROB measures were overturned in 34 percent, 54 percent, and 44 percent of cases, respectively. These results show some improvement over pre-2013, but are somewhat mixed across the PAs. The reader is referred to the tables in Appendix F for examination of individual PA results. Appendix F also includes lists of evaluated projects with overturned project types for PA review.

Table 5-3: PA vs. Evaluation Specified Project Type by Customer Agreement Date – All PAs

			PA-Specified Project Type								
			Add-on	**Capacity Expansion**	**Early Replacement**	**Major Renovation**	**New Construction**	**Natural Replacement**	**Replace on Burnout**	**System Optimization**	**Multiple**
Pre-2013 Customer Agreement Date											
Number of measures evaluated (n)			73								
Frequency of PA-Specified Measure Type (n)			31	1	7	0	15	0	8	5	6
Evaluation-Specified Project Type	Frequency of Measure-Level Obs.	(n)									
	Add-on	21	18	0	1	0	0	0	1	0	1
	Capacity Expansion	2	0	0	0	0	1	0	0	1	0
	Early Replacement	1	0	0	1	0	0	0	0	0	0
	Major Renovation	0	0	0	0	0	0	0	0	0	0
	New Construction	15	0	1	0	0	14	0	0	0	0
	Natural Replacement	5	1	0	1	0	0	0	2	0	1
	Replace on Burnout	8	2	0	2	0	0	0	4	0	0
	System Optimization	11	7	0	1	0	0	0	1	2	0
	Multiple	10	3	0	1	0	0	0	0	2	4
2013+ Customer Agreement Date											
Number of measures evaluated (n)			115								
Frequency of PA-Specified Measure Type (n)			47	1	13	0	27	1	16	4	6
Evaluation-Specified Project Type	Frequency of Measure-Level Obs.	(n)									
	Add-on	35	31	0	3	0	0	0	1	0	0
	Capacity Expansion	3	0	1	0	0	2	0	0	0	0
	Early Replacement	10	2	0	6	0	0	0	2	0	0
	Major Renovation	1	1	0	0	0	0	0	0	0	0
	New Construction	27	0	0	2	0	23	0	0	1	1
	Natural Replacement	6	0	0	0	0	1	1	2	1	1
	Replace on Burnout	16	3	0	1	0	0	0	9	0	3
	System Optimization	8	5	0	1	0	0	0	0	2	0
	Multiple	9	5	0	0	0	1	0	2	0	1

Figure 5-1 through Figure 5-4 show the distribution of ex-ante and ex-post defined project types for each PA. Across all PAs, a common theme is that PAs most frequently classify projects as add-on, early replacement, and replace on burnout (excluding new construction). These are also the most commonly overturned project types in the ex-post evaluation. Evaluators often overturn these commonly assigned project types to capacity expansion, natural replacement, and system optimization.

Figure 5-1: PG&E Distribution of Project Types – Ex-Ante and Ex-Post

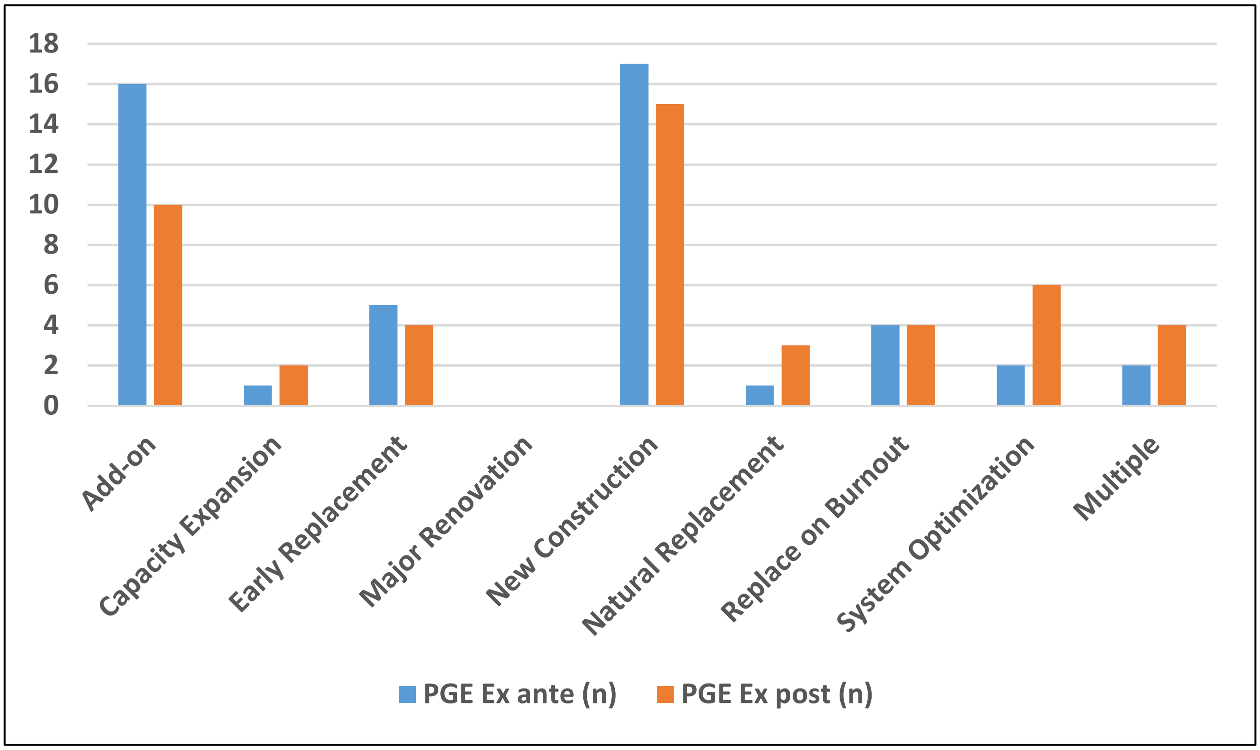


Figure 5-2: SCE Distribution of Project Types – Ex-Ante and Ex-Post

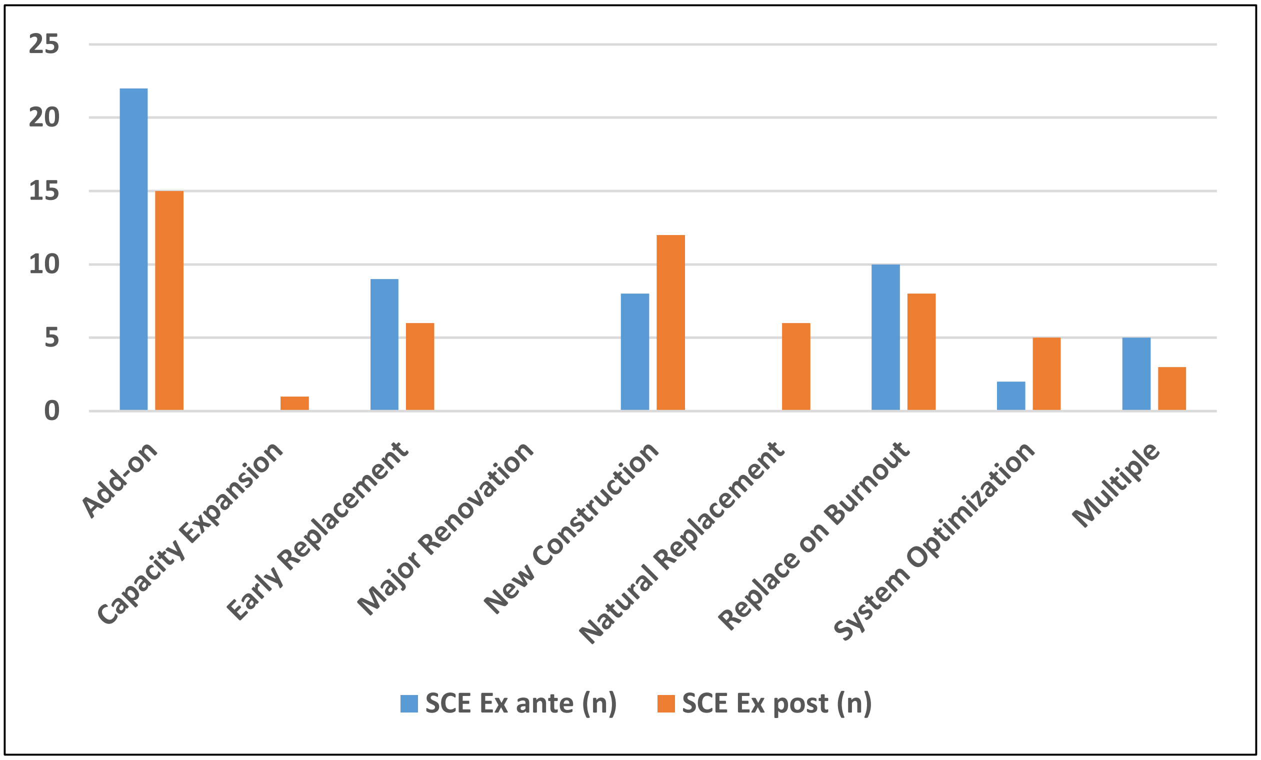


Figure 5-3: SCG Distribution of Project Types – Ex-Ante and Ex-Post

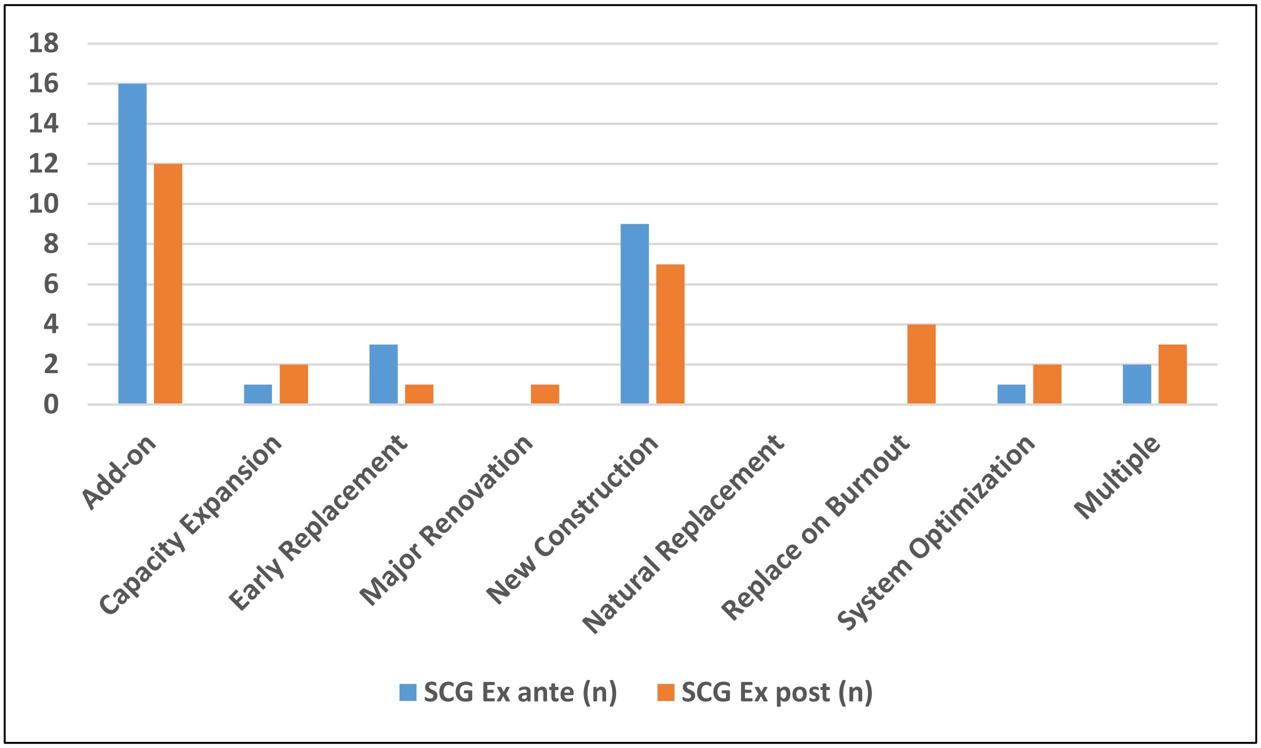
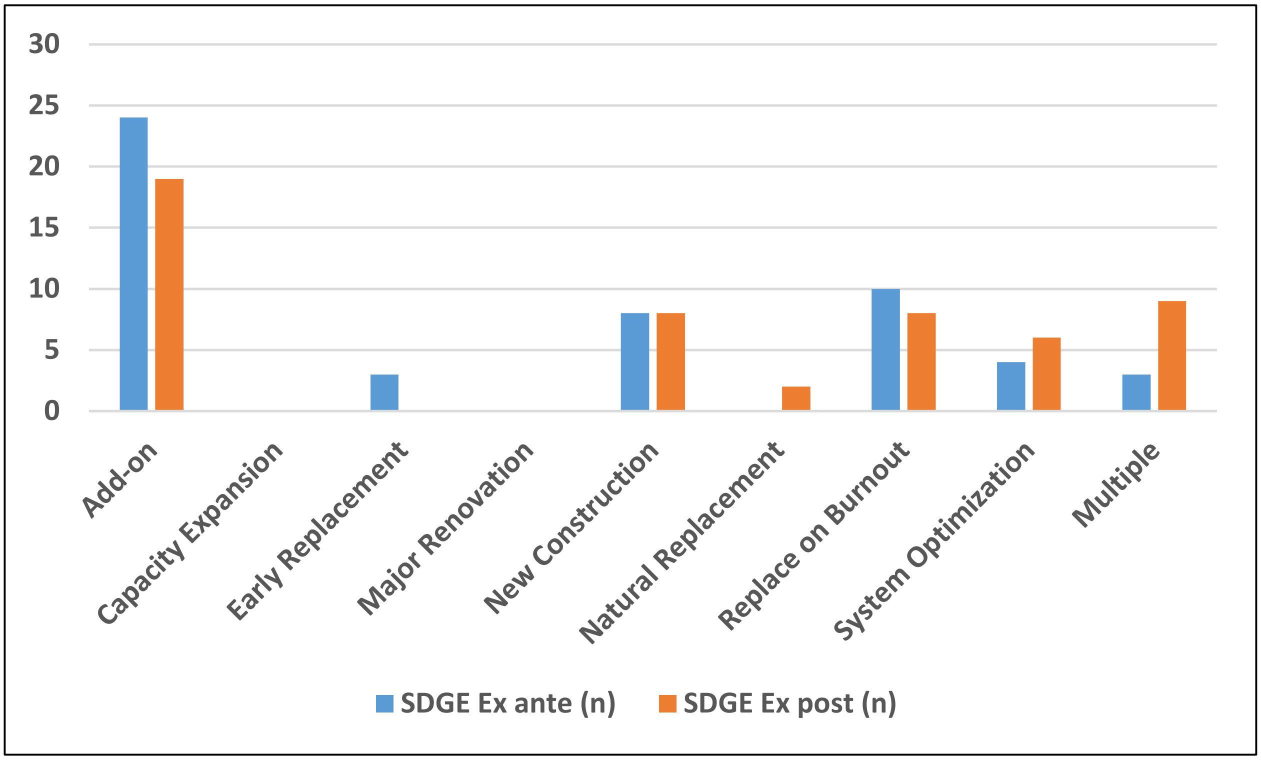


Figure 5-4: SDG&E Distribution of Project Types – Ex-Ante and Ex-Post



5.3.3 Project Baseline Assessment

As with project type, establishing the correct project baseline (existing equipment, Title 24, industry standard practice, etc.) is a critical first step in the project application process. The project baseline (and second baseline in the case of ER projects) forms the basis for accurately calculating first year and lifecycle measure savings.

Table 5-4 presents the frequency of ex-ante and ex-post agreement (and disagreement) on project baseline by PA and time period (pre-2013 customer agreement date versus 2013+). Across all PAs, there is a marginal decline in the accuracy of project baseline selection in the 2013+ period relative to the pre-2013 period (pre-2013: 74 percent agreement vs. 2013+: 68 percent agreement). Of the four PAs, only SCG showed an improvement in project baseline selection in the 2013+ period relative to pre-2013, and SDG&E exhibited a significant decline in baseline selection accuracy from 93 to 63 percent. SCE project baseline selection accuracy stayed constant over the two time periods at 63 percent, while PG&E accuracy declined from 87 to 70 percent.

In Chapter 3 it was shown that inappropriate baseline selection led to substantial downward adjustments to ex-ante gross MMBtu savings estimates. Increased efforts are needed from all PAs to ensure conformance with CPUC baseline policies including the examination and documentation of existing equipment RUL. For RUL this means carefully investigating and documenting the age, condition and functionality of existing equipment and operations, including the collection of measured data where warranted, and then subsequently using that information to establish proper baselines. Improvement in project type identification and documentation is one area of emphasis that will help promote proper baseline specification. Similar to what was noted above for improved eligibility treatment, greater levels of communication are needed with PA staff and contractors involved in implementing custom offerings, to ensure conformance CPUC baseline policies.

Table 5-4: Frequency of Ex-Ante and Ex-Post Agreement on Project Baseline by PA and Customer Agreement Date

PA	Customer Agreement Date	Project Baselines Matched	Project Baselines Overturned	Project Baselines % Matched
PG&E	Pre-2013	13	2	87%
	2013 +	23	10	70%
SCE	Pre-2013	20	12	63%
	2013 +	15	9	63%
SCG	Pre-2013	8	4	67%
	2013 +	16	4	80%
SDG&E	Pre-2013	13	1	93%
	2013 +	24	14	63%
All PAs	Pre-2013	54	19	74%
	2013 +	78	37	68%

Table 5-5 presents results detailing ex-ante versus ex-post project baseline designations for all PAs. The green shaded cells along the diagonal indicate the number of measures that showed agreement between the PA and ex-post evaluation. Values in the red shaded cells are measures where the project baseline was reassigned by the evaluator. As stated previously, the most commonly overturned project baseline was existing equipment. Improper use of an existing equipment baseline increased in the 2013+ period, leading largely to the trend noted above in Table 5-4 that accuracy of baseline specification for all PAs combined declined in the 2013+ period. Results for other baseline types are somewhat mixed across the PAs. The reader is referred to the tables in Appendix F for examination of individual PA results. Appendix F also includes lists of evaluated projects with overturned project baselines for PA review.

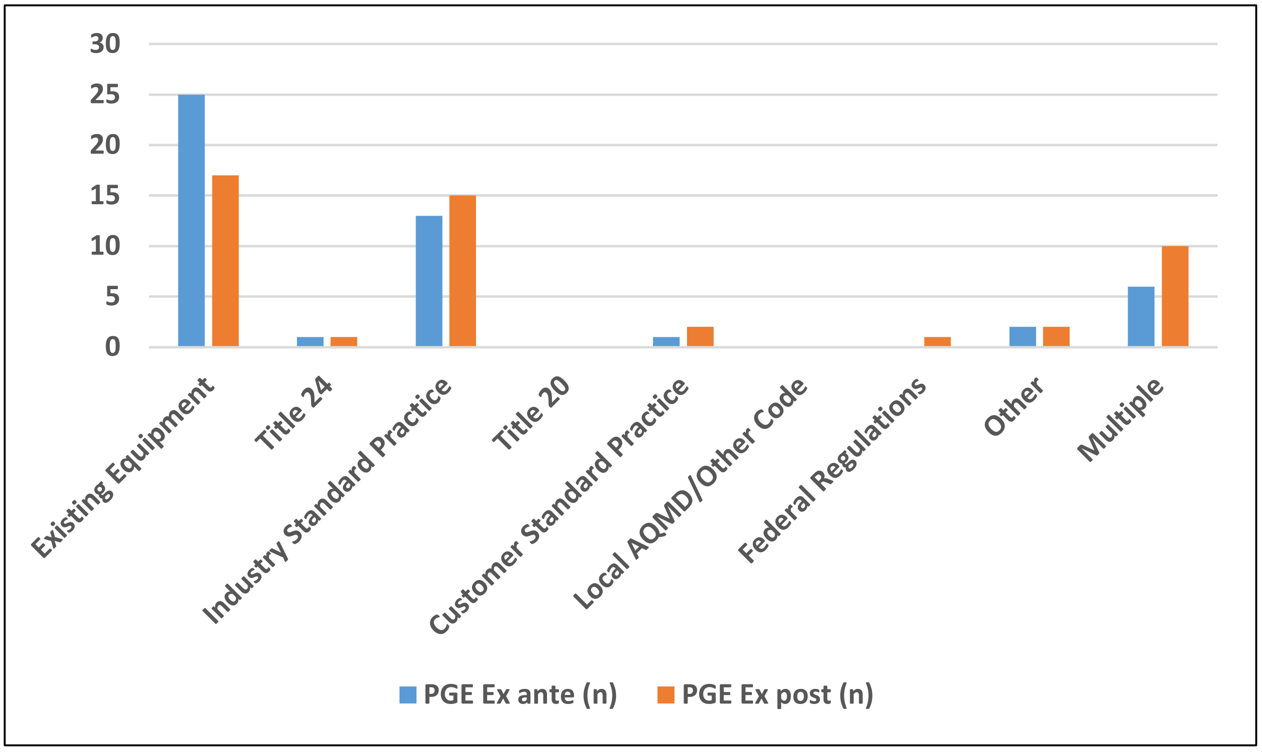
Table 5-5: PA vs. Evaluation Specified Project Baseline by Customer Agreement Date – All PAs

			All PA-Specified Project Baseline								
			Existing Equipment	Title 24	Industry Standard Practice	Title 20	Customer / Facility Std. Prac.	Local AQMD/ Other Code	Federal Regulations	Other	Multiple
Pre-2013 Customer Agreement Date											
Number of measures evaluated (n)			73								
Frequency of PA Specified Baseline (n)			50	5	5	2	0	0	0	4	7
Evaluation-Specified Project Baseline	Frequency of Measure-Level	(n)									
	Existing equipment	38	37	0	0	0	0	0	0	0	1
	Title 24	4	1	3	0	0	0	0	0	0	0
	Industry standard practice	12	7	0	4	0	0	0	0	1	0
	Title 20	1	0	0	0	1	0	0	0	0	0
	Customer/facility std. 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	1	0	1	0	0	0	0	0	0	0
	Other	4	0	0	1	0	0	0	0	3	0
	Multiple	13	5	1	0	1	0	0	0	0	6
2013+ Customer Agreement Date											
Number of measures evaluated (n)			115								
Frequency of PA-Specified Measure Type (n)			71	14	16	0	1	1	0	5	7
Evaluation-Specified Project Baseline	Frequency of Measure-Level	(n)									
	Existing equipment	46	44	0	0	0	0	1	0	0	1
	Title 24	14	2	11	0	0	0	0	0	1	0
	Industry standard practice	26	10	0	14	0	0	0	0	1	1
	Title 20	1	0	1	0	0	0	0	0	0	0
	Customer/facility std. practice	2	1	0	0	0	1	0	0	0	0
	Local AQMD/other code	1	1	0	0	0	0	0	0	0	0
	Federal regulations	3	1	2	0	0	0	0	0	0	0
	Other	7	3	0	1	0	0	0	0	3	0
	Multiple	15	9	0	1	0	0	0	0	0	5

Figure 5-5 through Figure 5-8 show the distribution of ex-ante and ex-post defined project baselines for each PA. Across all PAs, a common theme is that PAs most frequently establish existing equipment as the project baseline. That particular baseline was commonly overturned in the ex-post evaluation, and this is a common theme across custom evaluations going back at least a decade or more. Across both time periods, of the 121 measures with PA specified baselines of existing equipment, 40 were overturned, typically due to industry standard practice and "multiple."[69] In both cases, it is important that the PA fully establish and document project baselines for all component measures throughout the lifecycle of the project.

Existing equipment baselines were also overturned, to a lesser extent, due to federal/state codes and standards and customer / facility standard practice. 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.

Figure 5-5: PG&E Distribution of Project Baselines – Ex-Ante and Ex-Post



[69] An evaluator specified project baseline of "multiple" means either 1) it is an ER project in which the first and second baselines are different or 2) the "measure" is actually comprised of multiple component measures with different applicable baselines.

Figure 5-6: SCE Distribution of Project Baselines – Ex-Ante and Ex-Post

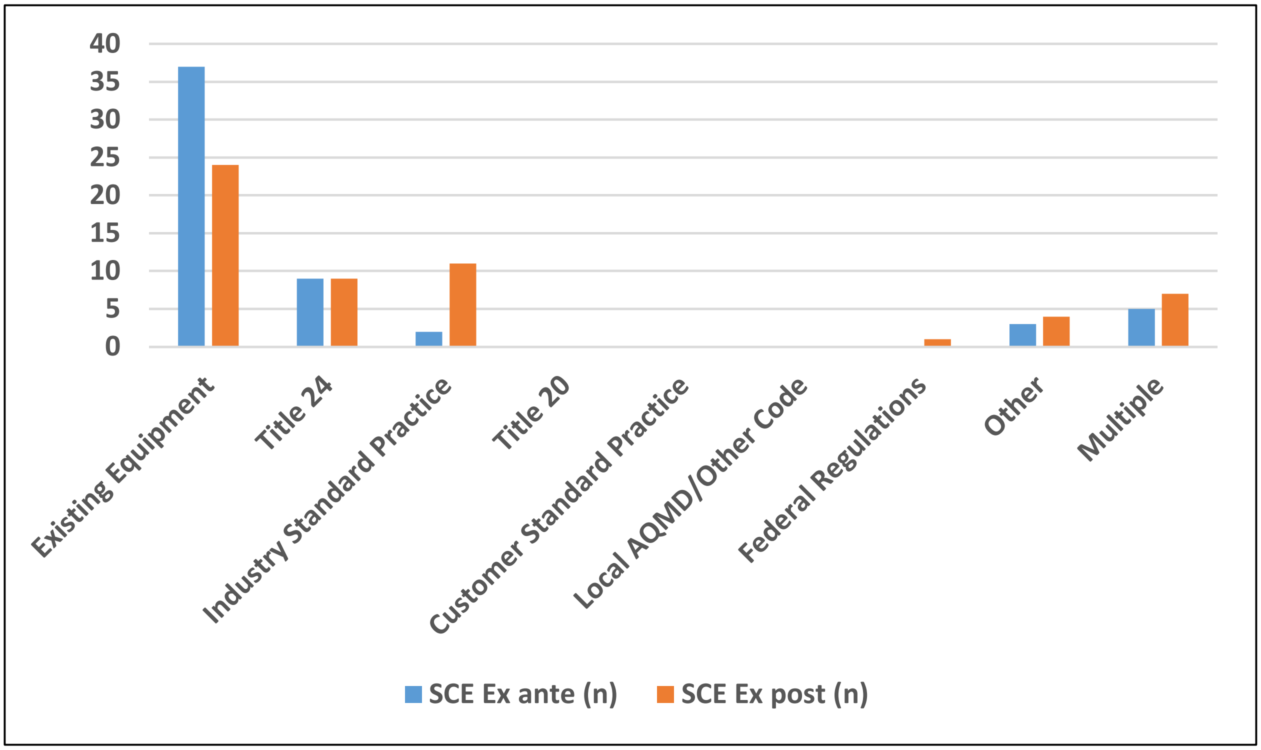


Figure 5-7: SCG Distribution of Project Baselines – Ex-Ante and Ex-Post

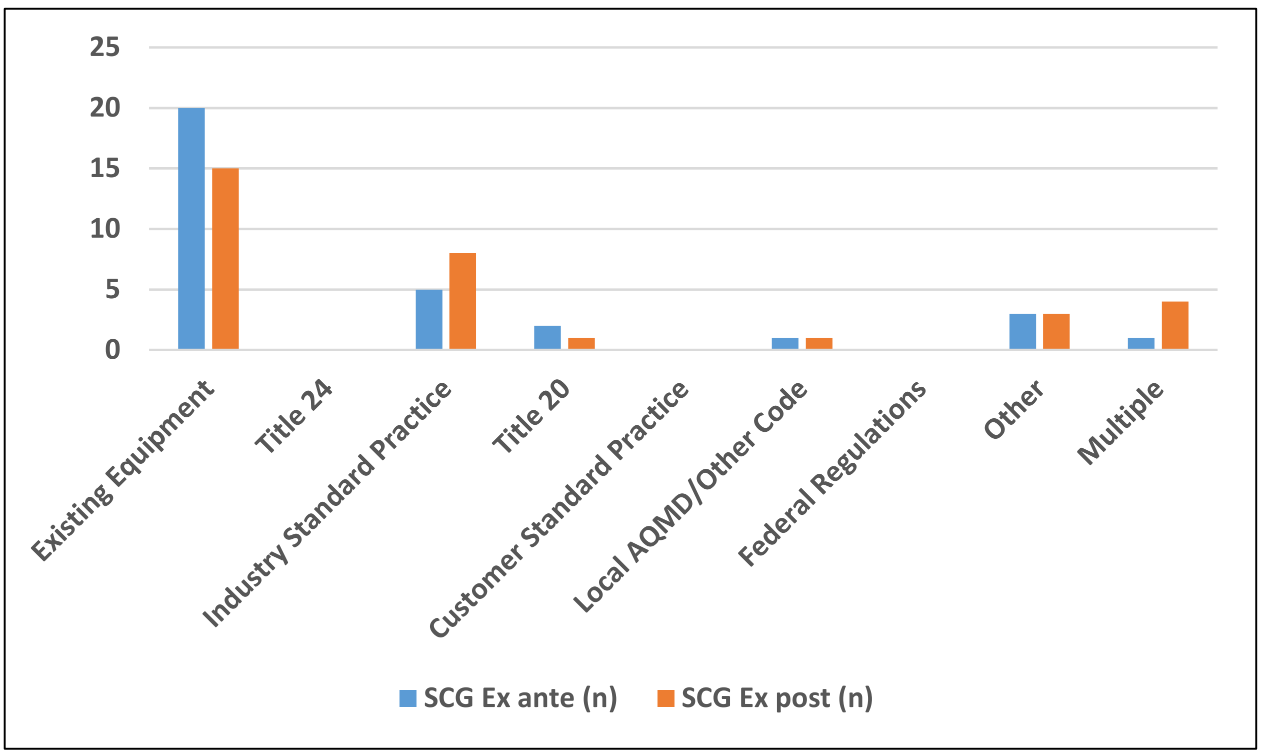
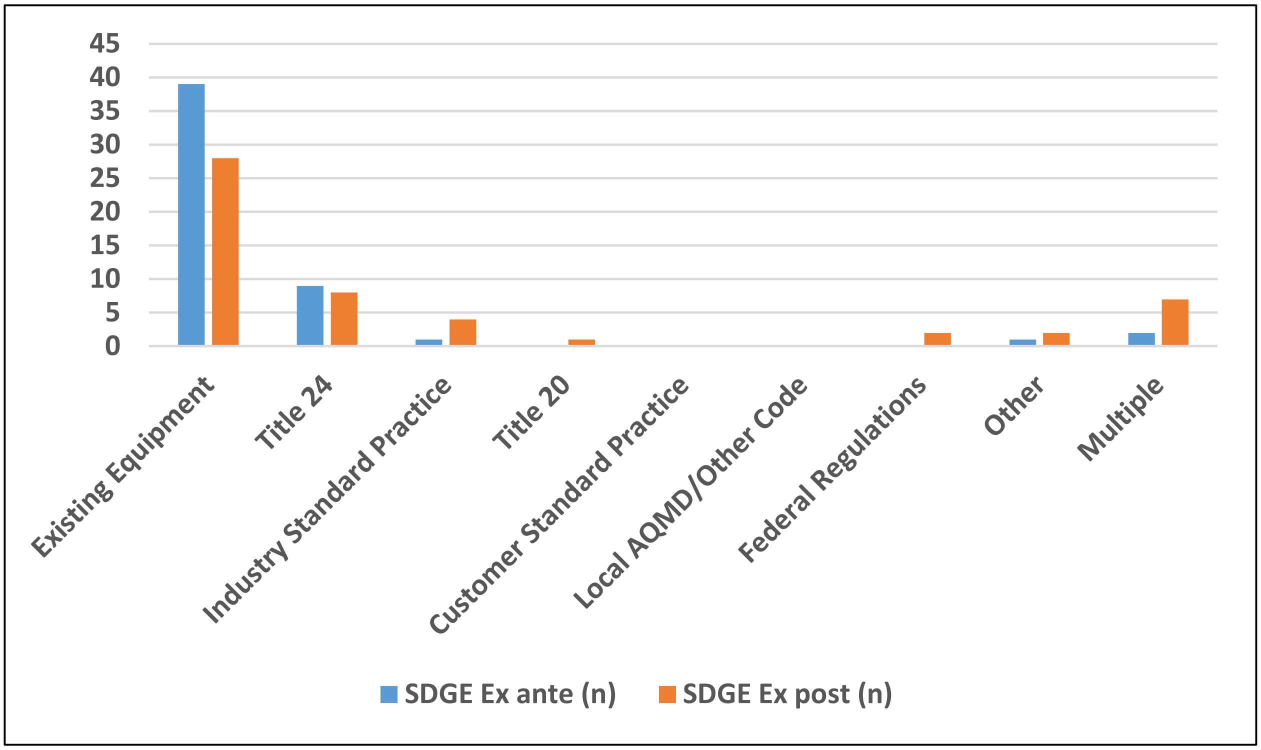


Figure 5-8: SDG&E Distribution of Project Baselines – Ex-Ante and Ex-Post



5.3.4 Project Baseline Ratings

Table 5-6 and Figure 5-9 (graphical representation) present a summary of the quality of the PA documentation that was used in establishing project baselines (Title 24, ISP, etc.). The Quality of Documentation score refers to the evaluator's rating of how well the PA documented their examination of the factors that determine baseline.[70] For example, ER, add-on, or system optimization projects should provide clear documentation of the age, condition, RUL, and the capability of performance through RUL of the existing equipment. Other relevant considerations include examination of facility and industry standard practices, applicable codes and standards, and maintenance records.

The Appropriateness of Baseline Determination score reflects 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. For example, documentation for an early replacement project should correctly establish the pre-existing system or equipment as the first baseline, accurately specify the second baseline, and include a narrative

[70] Recall that appropriate project type determination is also a critical factor that should be documented and subsequently incorporated into establishment of baseline for a given project.

for the second baseline assignment. Low baseline appropriateness scores generally correspond to overturned baselines demonstrated above in Table 5-5.

The Quality of Baseline Description rating scores PAs on the accuracy and completeness of their baseline description. The baseline description should include a description of the correct baseline equipment and its efficiency. Again, for ER projects, both baselines should be accurate and adequately described, including descriptions of the EUL and RUL periods.

Table 5-6: Project Baseline Documentation Quality by PA and Customer Agreement Date

Parameter Examined			PA Project Baseline 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
Project Baseline	Quality of Documentation Rating	n	15	32	12	14	33	24	20	38
		Mean	2.6	2.5	2.58	2.43	2.48	2.33	2.3	2.05
		Median	3	3	3	3	3	2	2	2
	Appropriateness of Baseline Determination Rating	n	15	30	12	14	33	23	20	38
		Mean	2.8	2.47	2.25	2.64	2.58	2.35	2.5	2.13
		Median	3	3	2.5	3	3	3	3	2
	Quality of Baseline Description Rating	n	15	31	12	14	33	22	20	35
		Mean	2.87	2.52	2.67	2.71	2.55	2.59	2.6	2.2
		Median	3	3	3	3	3	3	3	2

Figure 5-9: Mean Project Baseline Documentation Quality Ratings by PA and Customer Agreement Date

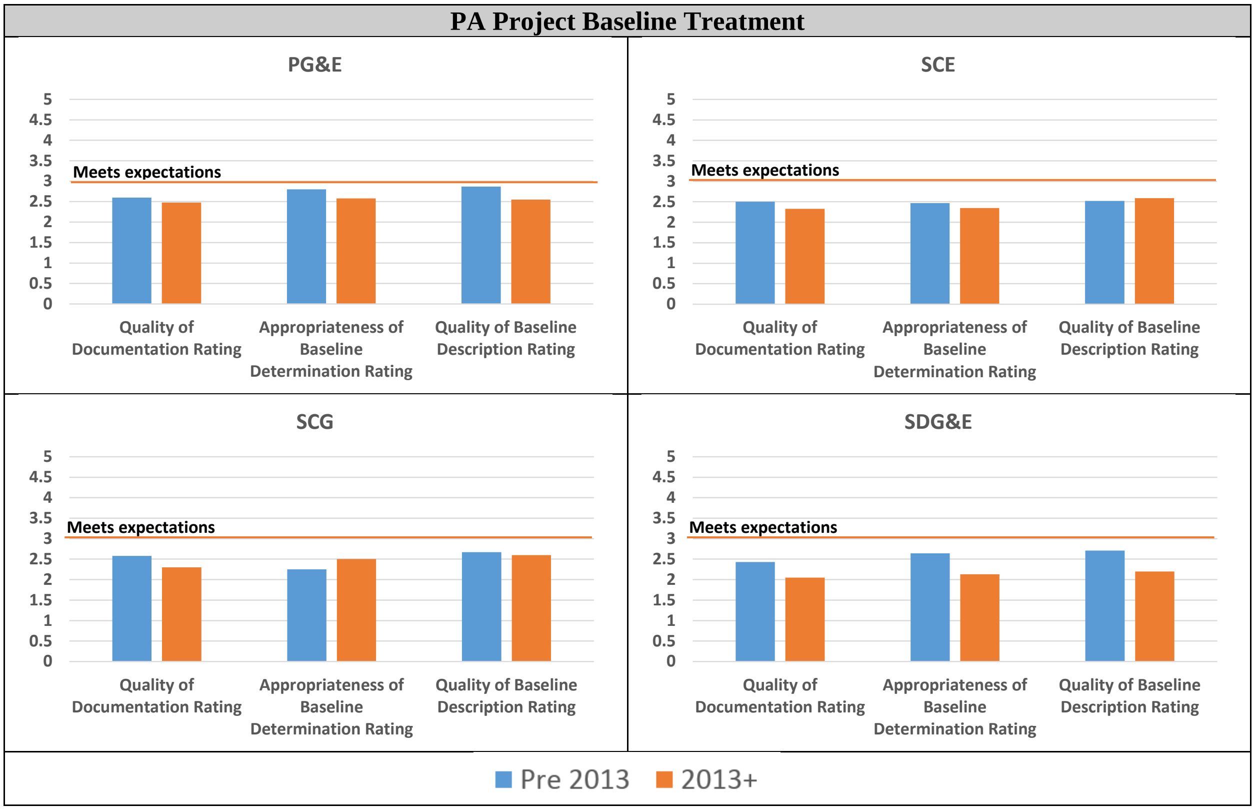


Table 5-6 and Figure 5-9 show that on average, none of the PAs "met expectations" (i.e. mean score of 3) in either the pre-2013 or 2013+ time period. This is true for all three metrics. Scores range from 2.25 to 2.8 in the pre-2013 period and from 2.05 to 2.6 in the 2013+ period. Nearly all scores for all PAs and all three metrics were lower in the 2013+ period (except SCE baseline description and SCG appropriateness metrics). It should be noted that the majority of scores in both the pre-2013 and 2013+ periods have medians of 3, which in this case reflects that the majority of projects are meeting expectations for these metrics. However, the overturned project types and project baselines described in the previous sections are associated with low scores, which brings the averages down. The scores presented here are expected to become more useful to the reader when trends are explored as part of the 2015 evaluation, allowing the reader to better examine change with time.

In instances where the documentation quality scores did not meet expectations (scores of 1 or 2), the evaluation most often cited the condition of existing equipment, the age of the existing equipment, the capability of existing/baseline equipment to meet facility service needs, normal facility practices, applicable codes and standards, the EUL of the equipment, and industry standard practice when correctly establishing the baseline. The PAs should strive to thoroughly examine and document each of these common factors (among others) when establishing project type and baseline. As noted above, these cases of inadequate PA documentation 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, evaluators more often reviewed ISP and applicable codes and standards, which were also common reasons for overturned project baselines.

To enhance PA documentation of project type and baseline, as well as an array of parameters and engineering conclusions, a statewide form would be useful for recording critical information used to make PA choices, including triangulation where multiple data points contribute to a given conclusion. The evaluation uses PPA elements in the final site report form to record such information, and the PAs should examine this form and consider augmenting it for the purposes of improving documentation for all projects.

5.3.5 EUL Assessment

Table 5-7 provides a comparison of the EUL values that were documented in the PA tracking data, project application files, and the ex-post evaluation. EUL was populated in the tracking data for nearly 100 percent of measures (two PG&E measures and one SDG&E measure did not have EULs populated in the 2013+ period). By comparison, EUL was poorly documented in the project application files, ranging from 13 percent to 83 percent of measures in the pre-2013 period and from 33 percent to 75 percent in the 2013+ period. While these scores are generally higher than those in the 2013 evaluation (where the highest rate of documentation was 64 percent), all PAs are still providing insufficient EUL data in the project files.

Table 5-7: 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
Summary of Evaluation EUL Treatment								
Number of Measures Assessed	15	32	12	14	33	24	20	38
Number of Measures with PA Tracking System EUL Populated (n)	15	32	12	14	31	24	20	37
Percent of Measures with PA Tracking System EUL Populated	100%	100%	100%	100%	94%	100%	100%	97%
Number of Measures with PA EUL Documented in the Project Application Files	2	23	10	9	11	18	15	27
Percent of Measures with PA EUL Documented in the Project Application Files	13%	72%	83%	64%	33%	75%	75%	71%
Mean PA EUL Documentation Score	1.5	2.3	2.4	2.0	1.7	2.4	2.3	2.1
Median PA EUL Documentation Score	1.0	3.0	2.0	2.0	2.0	3.0	2.0	2.0
Summary of EUL Differences								
Number of Measures with Evaluation EUL Different Than PA Tracking EUL (n)	9	13	5	9	24	6	6	18
Percent of Measures with Evaluation EUL Different Than PA Tracking EUL	60%	41%	42%	64%	73%	25%	30%	47%
Mean Evaluation EUL (where differences exist)	12.3	8.3	10.2	7.2	11.8	9.8	9.2	9.8
Median Evaluation EUL (where differences exist)	10.0	8.0	8.0	5.0	10.5	10.0	8.3	8.4
Mean PA Tracking System EUL (where differences exist)	14.7	13.9	13.4	14.4	11.9	14.7	15.2	13.8
Median PA Tracking System EUL (where differences exist)	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0

PA documentation of EUL in the project application files is summarized as follows:

- PG&E was deficient in their documentation of EUL in the project application files (pre-2013: 13 percent; 2013+; 33 percent). PG&E's low mean EUL documentation scores of 1.5 and 1.7 largely reflect the missing EUL documentation and, to a lesser extent, EUL's that were documented but incorrect (5 of 13 or 38 percent).
- SCE was relatively consistent in their documentation of EUL in the project application files between the two periods, with 72 percent and 75 percent of records documented. For records that had EUL documentation, 34 percent were incorrect. These factors led to EUL documentation scores of 2.3 and 2.4 for the two time periods.
- SCG's rate of EUL documentation was relatively high at 83 percent in the pre-2013 period and 75 percent in 2013+. Overall SCG's results were very similar to those reported for SCE above, with 32 percent of documented records having incorrect conclusions regarding EUL.
- SDG&E's rate of documentation of EUL in the project application files improved from 64 percent in the pre-2013 period to 71 percent in the 2013+ period. However, the rate of incorrect EUL documentation for those records was relatively high, at 39 percent. This led to low mean documentation scores of 2.0 and 2.1.

Of the 115 measures that had EUL documented in the project application files, 62 (54 percent) were sourced from DEER. No source was provided for 32 of the EULs documented in the project applications. Similarly, DEER was the primary source of evaluation-sourced EULs; 113 out of 188 evaluation EULs were sourced from DEER (60 percent).

Across all PAs the mean evaluation-sourced EULs were lower than the mean tracking system EULs. For the 48 percent of measures that had different PA tracking and evaluation-sourced EULs (a total of 90 out of 188 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 SCG. In the 2013+ period, PG&E's tracking system EULs matched the evaluation EULs almost exactly. Nine of their 33 records did not match evaluation conclusions regarding EUL, but those differences were minor. The mean differences between the evaluation-sourced and tracking EULs for each PA and each application period are as follows:

- PGE - pre-2013: -2.4, 2013+: -0.01;
- SCE - pre-2013: -5.5; 2013+: -4.9;
- SCG - pre-2013: -3.2; 2013+: -5.9;
- SDG&E - pre-2013: -7.3; 2013+: -4.0.

When the EUL differences noted above are combined across all PAs and time periods, the simple average evaluation EUL is approximately 4.2 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 (for all PAs except PG&E) and this EUL difference was *a key factor* driving down the LC GRR results. It should be noted that for all 81 PG&E measures, the mean difference between the tracking system EUL and evaluation derived EUL was 0.08, which helps to explain why PG&E is the only PA that shows lower FY GRRs than LC GRRs. Evaluation-derived and applied weights also contributed to this PG&E difference, as explained in Chapter 3.

5.3.6 Calculations Assessment

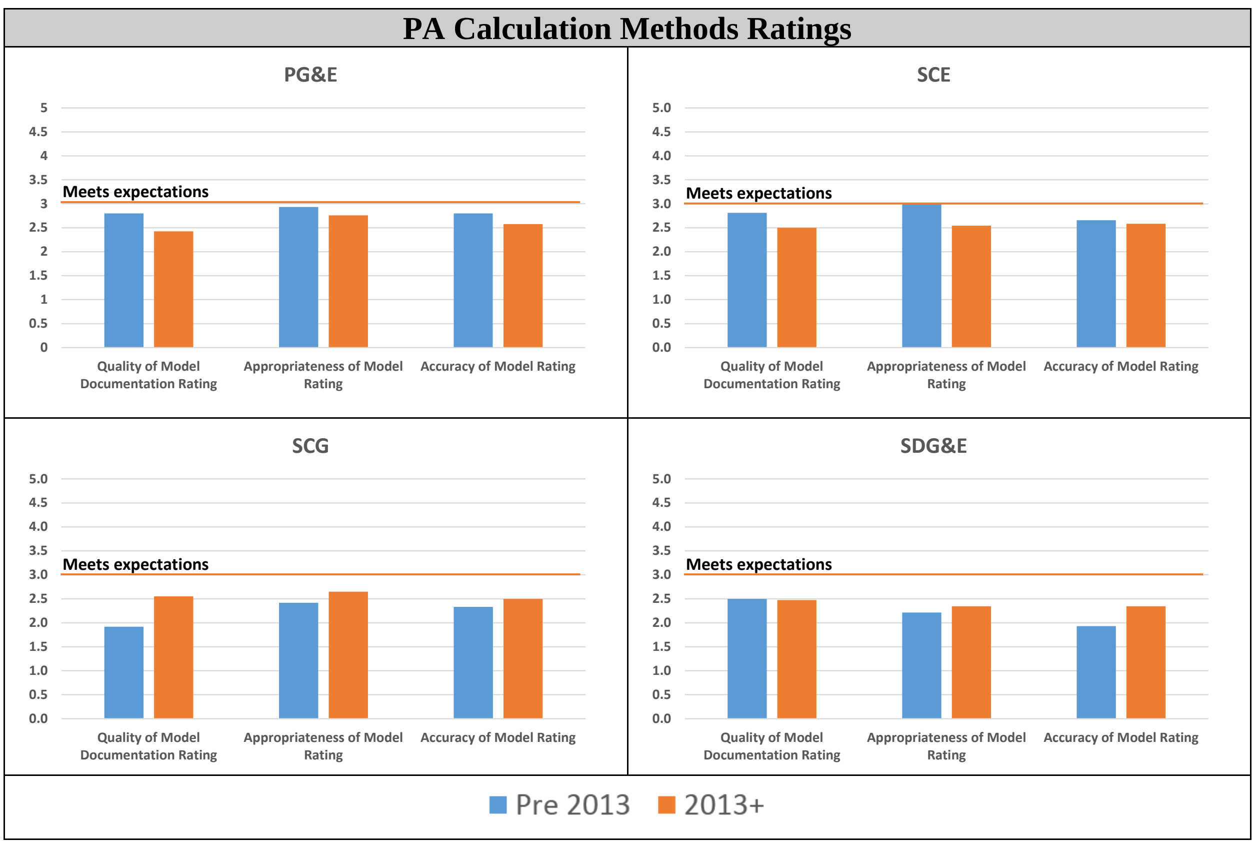
Table 5-8 and Figure 5-10 (graphical representation), below, provide an assessment of the documentation quality, appropriateness, and the accuracy of the PA models in determining measure savings. The Quality of Documentation score reflects the degree to which the PA calculation model is clearly documented for *both* the pre- and post-installation conditions. Key parameters and parameter relationships should be highlighted, and the model itself should be unlocked, in an accessible format, and include any relevant input or output files.

The Appropriateness of Model Score quantifies whether the PA calculation model is suitable for the project and whether it accounts for key parameters that could impact savings such as weather, production, or seasonal adjustments. The Accuracy of Model score rates the extent to which the PA calculation model uses site-specific values and reliable typical input values (such as, flow rates, pressures, temperatures, weather data or production data).

Table 5-8: Calculations Methods Assessment by PA and Customer Agreement Date

Parameter Examined			PA Calculation Methods 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**
Calculation Methods	Quality of Model Documentation	n	15	32	12	14	33	24	20	38
		Mean	2.8	2.8	1.9	2.5	2.4	2.5	2.6	2.5
		Median	3.0	3.0	2.0	3.0	3.0	3.0	3.0	3.0
	Appropriateness of Model	n	15	32	12	14	33	24	20	38
		Mean	2.9	3.0	2.4	2.2	2.8	2.5	2.7	2.3
		Median	3.0	3.0	2.5	2.0	3.0	3.0	3.0	2.0
	Accuracy of Model	n	15	32	12	14	33	24	20	38
		Mean	2.8	2.7	2.3	1.9	2.6	2.6	2.5	2.3
		Median	3.0	3.0	2.5	2.0	3.0	3.0	3.0	2.0
Assessment of Evaluation Use of PA Inputs										
Number of Measures Assessed (n)			14	29	12	13	30	23	17	37
Evaluation used a different model			14%	10%	42%	15%	27%	26%	24%	27%
Evaluation used a similar model			43%	34%	58%	38%	50%	26%	65%	49%
Evaluation adjusted the PA model			43%	55%	0%	46%	23%	48%	12%	24%

Figure 5-10: Calculations Methods Assessment by PA and Customer Agreement Date



For PG&E and SCE, scores for these three metrics ranged from 2.4 to 3.0 across the two time periods. This indicates that the calculation methods for these PAs were generally appropriate, accurate, and well documented. However, average scores for the 2013+ period were lower for all three metrics than in pre-2013 applications.

Calculation methods scores for SCG and SDG&E ranged from 1.9 to 2.5 in the pre-2013 period and 2.3 to 2.7 in the 2013+ period. Scores for all metrics were higher in the 2013+ period, with the exception of the documentation quality score for SDG&E.

Finally, Table 5-8 shows that in the pre-2013 period, the evaluator only used the PA model (or similar model) for 34 percent to 58 percent of measures. In the 2013+ period, this figure ranged from 26 to 65 percent, and increased for all PAs except SCE. In all other cases, the evaluator used an entirely different model or deemed it necessary to make adjustments to the PA models. Even in the cases where PAs received high calculation methods scores (e.g. PG&E and SCE in the pre-2013 period), evaluators commonly had to adjust PA models or re-model measures to properly estimate savings.

There is room for improvements to PA impact methods and models, through incorporation of industry best practices, careful review of evaluation approaches/differences and continued participation in the ex-ante review process. Due diligence is also warranted for the purposes of ensuring that the PAs adhere to CPUC impact estimation policies and requirements. PA technical staff reviews of savings estimates and calculations should be thorough and conducted prior to finalization of incentives and savings claims.

5.3.7 Inputs and Assumptions Assessment

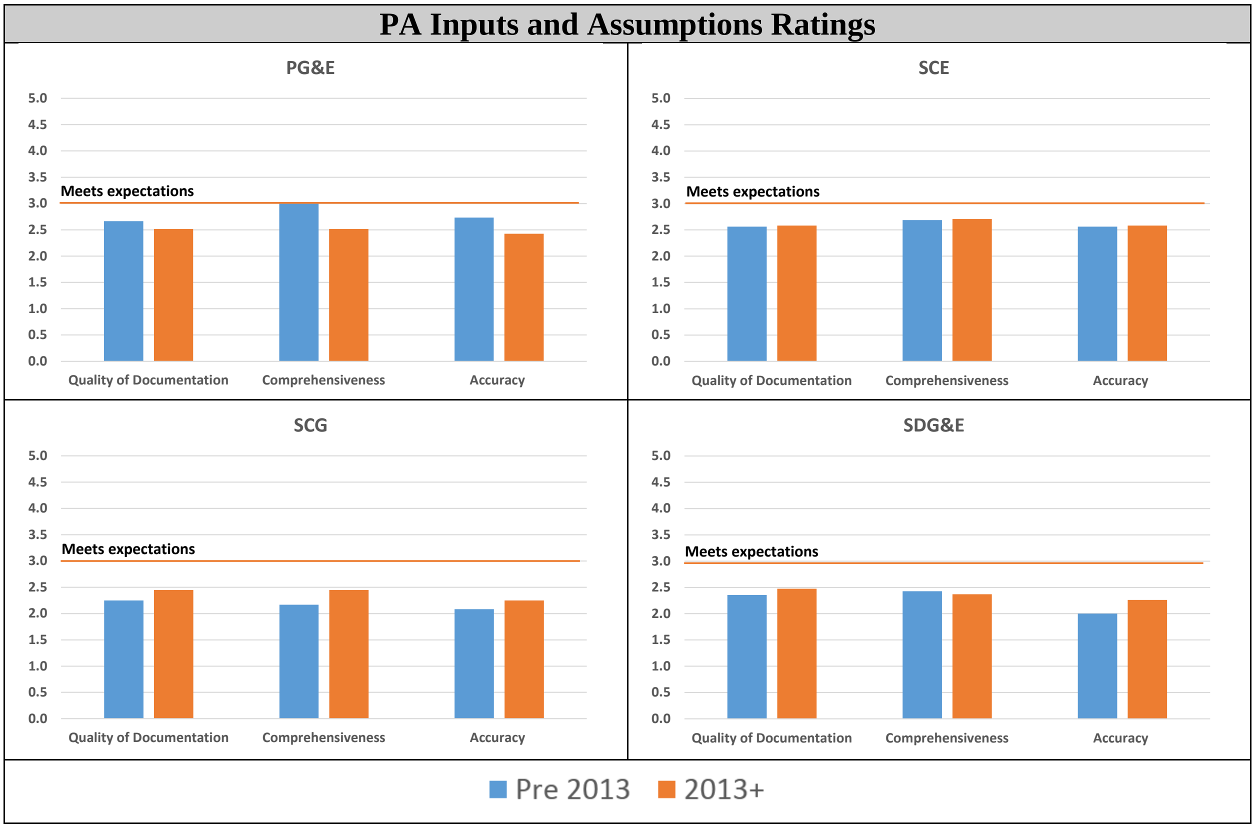
Table 5-9 and Figure 5-11 (graphical representation) summarizes the documentation quality, comprehensiveness, and accuracy ratings for the PAs' calculation method inputs and assumptions and provides an assessment of the evaluation team's use of the PAs' inputs and assumptions. The Quality of Documentation score rates the degree to which PA inputs and assumptions are accompanied by clearly documented sources. In order to receive a score of "3" (meets expectations), the PA must provide supporting sources for the most important inputs and assumptions (those parameters having a high impact on savings).

The Comprehensiveness score reflects the extent to which the PA included *all* relevant inputs and assumptions in the model. A score of "3" here indicates that the calculation model includes the most relevant inputs and assumptions (e.g., load factor, efficiency, flow, power factor, weather, production or seasonal adjustments performed). Finally, the Accuracy score quantifies the correctness of the most relevant inputs and assumptions. All relevant inputs and assumptions must be deemed accurate by the evaluation engineer in order to receive a score of three.

Table 5-9: Inputs and Assumptions Assessment by PA and Customer Agreement Date

Parameter Examined			PA Inputs and Assumptions 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
Inputs and Assumptions	Quality of Documentation	n	15	32	12	14	33	24	20	38
		Mean	2.7	2.6	2.3	2.4	2.5	2.6	2.5	2.5
		Median	3.0	3.0	2.0	2.5	3.0	3.0	3.0	3.0
	Comprehensiveness	n	15	32	12	14	33	24	20	38
		Mean	3.0	2.7	2.2	2.4	2.5	2.7	2.5	2.4
		Median	3.0	3.0	2.0	2.5	3.0	3.0	3.0	2.5
	Accuracy	n	15	32	12	14	33	24	20	38
		Mean	2.7	2.6	2.1	2.0	2.4	2.6	2.3	2.3
		Median	3.0	3.0	2.0	2.0	3.0	3.0	2.0	2.0
Assessment of Evaluation Use of PA Inputs										
Number of Measures Assessed (n)			15	28	12	13	32	24	19	37
Evaluation used a different inputs			27%	7%	33%	8%	22%	29%	26%	16%
Evaluation used a similar inputs			40%	39%	42%	69%	47%	29%	53%	59%
Evaluation adjusted the PA inputs			33%	54%	25%	23%	31%	42%	21%	24%
Assessment of Evaluation Use of PA Assumptions										
Number of Measures Assessed (n)			15	27	8	13	31	23	16	35
Evaluation used a different assumptions			20%	19%	50%	38%	29%	35%	50%	20%
Evaluation used a similar assumptions			47%	59%	38%	38%	29%	39%	38%	54%
Evaluation adjusted the PA assumptions			33%	22%	13%	23%	42%	26%	13%	26%

Figure 5-11: Inputs and Assumptions Assessment by PA and Customer Agreement Date



As shown in the table, the mean documentation, comprehensiveness, and accuracy ratings for each PA are between 2.0 and 3.0, indicating that, on average, the PAs fell short of minimum expectations (with the exception of PG&E's pre-2013 comprehensiveness score). PA specific results are summarized as follows:

- PG&E's average ratings for the three inputs and assumptions evaluation categories declined from a range of 2.7 to 3.0 in the pre-2013 period to 2.4 to 2.5 in the 2013+ period.
- SCE's average ratings for the three inputs and assumptions evaluation categories stayed steady in both periods, with all scores in the 2.6 to 2.7 range.
- SCG's average input and assumptions scores showed some improvement in the 2013+ period, but all scores remained at 2.5 or below.
- All of SDG&E's average ratings for the three inputs and assumptions categories were also 2.5 or less. However, there was marginal improvement in the documentation quality and accuracy scores in the 2013+ period.

The evaluation team also documented the use of different, similar, or adjusted inputs and assumptions compared to those used in the PAs' calculation methods. Across all PAs and time periods, evaluators used similar inputs to the PAs (or the same) for 29 percent to 69 percent of records and similar assumptions for 29 percent to 59 percent of records. This shows that evaluators typically needed to adjust or entirely replace PA defined inputs and assumptions. There are no clear trends in the pre-2013 to 2013+ periods.

Project inputs and assumptions should incorporate the use of pre- and post-installation data and information where possible. This way savings calculations can be based on actual equipment use schedules and reflect post-installation operating parameters such as flow rates, temperatures, set points, system pressures, production rates and power measurements.

5.3.8 Incremental Cost Assessment

This assessment only examines the first order question of whether or not incremental costs are documented in the project application files. Incremental cost ratings were only assessed where applicable project types were assigned by the PA (ER, ROB, NR, NC, and capacity expansion). Table 5-10 presents these results.

Table 5-10: Incremental Cost Documentation by PA and Customer Agreement Date

Parameter Examined	PA Incremental Cost Treatment							
	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* (n)	9	14	6	2	19	13	7	19
Number of Measures with Incremental Cost Populated (n)	7	9	5	1	11	8	7	11
Percent of Measures with Incremental Cost Populated	78%	64%	83%	50%	58%	62%	100%	58%

* Measures examined for PA incremental cost treatment includes only cases where incremental cost is applicable. Determination of incremental cost applicability is based on the PA conclusion of project type being early replacement, replace on burnout, natural replacement, new construction or capacity expansion. Incremental project cost is not relevant for other project types, including add-on and system optimization.

Small sample sizes make it difficult to draw conclusions by custom agreement date-based period, so results are only discussed here at the PA level.

- SCG documented incremental cost where appropriate to do so in 12 out of 13 cases that were examined (92 percent). It was also noted in Chapter 3 that SCG consistently reports this type of information in the tracking system, and ensures that fields are populated when appropriate to do so. This includes following CPUC guidance and decisions, and linking together reporting of project type, baseline, first period savings, second period savings, RUL and EUL. The other PAs should strive to report with equal accuracy.

- The other PAs documented incremental cost where appropriate to do so much less frequently, ranging from 57 percent to 64 percent.

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. While the need for PA attention to each recommendation varies based on the results of this evaluation, in general all recommendations apply to all PAs to some degree.

Many of the findings and recommendations presented in this chapter are the same or similar to those found in the 2013 custom impact evaluation report, as well as previous California custom impact evaluations. This is because findings and issues identified in the past still persist in 2014, and as a result suggestions for improving custom program implementation have not changed substantially. It is notable, however, that progress is being made to address previous findings and recommendations, including discussions with the CPUC and ex-ante review teams, improvements that are being made to within-PA processes, as well as coordinated activities across the PAs. However, while filing their Response to Recommendations (RTRs) for the 2013 report, the PAs did not explicitly agree to implement some recommendations pending further discussion with the CPUC and the ex-ante team. For example, applicability of DEER required methods and inputs for DEER-like measures, use of a payback floor to reduce free-ridership levels, appropriate level and duration for M&V, use of statewide standardized impact calculation tools, post-installation true-up with stable operation, etc. have been identified as areas for collaborative discussions. For this reason this chapter identifies all unique findings and recommendations that have not previously appeared by marking those paragraphs with a double asterisk (**).

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 Other Findings and Recommendations

The chapter begins with an examination of recent trends in evaluation-based gross impact realization rate (GRR) results.

**As summarized in Table 6-1 (and Chapter 3), ex-post MMBtu lifecycle gross impact realization rates (LC GRRs) range by PA from 0.49 to 0.63.[71] These MMBtu GRRs are not statistically different from either the 2010-2012 or 2013 evaluations. However, relative to the 2013 custom impact evaluation results PG&E and SCG average LC GRRs *decreased* by about two percent and 19 percent, respectively on a combined MMBtu basis. LC GRR results for SCE *increased* by 31 percent and SDG&E *increased* 28 percent.

Table 6-1: 2010-2012, 2013, and 2014 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	2013 Mean LC Gross Realization Rate	2013 90% Confidence Interval	2014 Mean LC Gross Realization Rate	2014 90% Confidence Interval
PG&E						
MMBtu*	0.63	0.57 to 0.69	0.63	0.57 to 0.70	0.62	0.50 to 0.73
kW	0.46	0.35 to 0.58	0.44	0.28 to 0.61	0.74	0.34 to 1.14
SCE						
MMBtu	0.61	0.51 to 0.71	0.44	0.34 to 0.54	0.58	0.44 to 0.71
kW	0.57	0.47 to 0.67	0.52	0.43 to 0.62	0.46	0.34 to 0.58
SDGE						
MMBtu*	0.57	0.47 to 0.67	0.49	0.40 to 0.59	0.63	0.57 to 0.70
kW	0.82	0.46 to 1.17	0.76	0.57 to 0.95	0.63	0.54 to 0.71
SCG						
MMBtu	0.64	0.54 to 0.75	0.60	0.48 to 0.72	0.49	0.36 to 0.62

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

**Net-to-gross ratios (Table 6-2), ranging from 0.46 to 0.62, are lower for all PAs than in the 2013 evaluation, ranging from a five percent decrease for SCG to a 19 percent decrease for SCE. However, 2014 NTG values are similar to those observed in 2010-2012, with the exception of SCG, which had a higher NTG in both 2013 and 2014.

[71] Ex-post gross impact results were also developed in this evaluation for MMBtu first year realization rates (FY GRRs), which range by PA from 0.58 to 0.73. Relative to the 2013 custom impact evaluation results PG&E, SDG&E, and SCG FY MMBtu GRRs *decreased* by about 20 percent, three percent, and 16 percent, respectively. FY GRR results *increased* for SCE by about 17 percent. It is notable that FY GRRs are an indication of performance in conducting ex-ante engineering-based savings estimates and associated PA processes, whereas LC GRRs are an indication of performance in a combination of engineering-based savings estimation and EUL and early retirement (ER) treatment (including associated RUL and EUL considerations). LC MMBtu GRRs were lower than the corresponding FY GRRs for SCE, SDG&E, and SCG.

Table 6-2: Comparison of 2010-12, and 2013, and 2014 Weighted MMBtu* NTGR Results

Energy Metric	2010-2012 Mean NTGR	2010-2012 90% Confidence Interval	2013 Mean NTGR	2013 90% Confidence Interval	2014 Mean NTGR	2014 90% Confidence Interval
PG&E						
MMBtu*	0.51	0.49 to 0.52	0.55	0.52 to 0.59	0.51	0.47 to 0.54
SCE						
MMBtu*	0.49	0.47 to 0.50	0.57	0.52 to 0.61	0.46	0.42 to 0.49
SDGE						
MMBtu*	0.48	0.46 to 0.50	0.59	0.55 to 0.64	0.51	0.47 to 0.56
SCG						
MMBtu*	0.49	0.40 to 0.58	0.66	0.59 to 0.73	0.62	0.60 to 0.65

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

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,[72] including a review of evaluation EUL conclusions/rationale in an effort to improve EUL claims and LC GRR results,
  - 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 improve quality control, PAs should increase due diligence on accuracy, comprehensiveness and documentation in project application files. SCG use of documentation and consistency in reporting improved in the 2013+ period, and SCG was also successful in transferring documented conditions from the application forms and reports into the tracking system in a consistent manner. PG&E also performed well in terms of documentation and tracking.

[72] It is notable that the evaluation estimate of EUL differed from the PAs estimate 48 percent of the time. For those instances the evaluation-derived average EUL was smaller than the ex-ante average EUL by nearly 4 years, representing a 26 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 *a key factor* driving down the LC GRR results.

- 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 PA 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:

- Underperforming Projects
- Project Operating Conditions
- Project Baseline Specification
- Project Calculation Methods

6.1.1 Underperforming Projects

All PAs had projects with negative and/or zero GRRs, and these served to lower the weighted realization rate considerably. Out of 150 M&V points, 22 projects, or 15 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 13 of the cases were due principally to one of two factors – inappropriate baseline or ineligible measures. While all PAs had negative and/or zero GRRs associated with projects that were determined to be ineligible, the cases related to baseline determination were applicable to just PG&E and SCE. For SCE eligibility accounted for 32 percent of all downward adjustments to ex-ante claims.

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 eligibility 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 should focus on the following issues identified in Chapter 3: improved attention to ISP determinations and their effective dates, assurance that impacts are realized on the grid where on-site generation is present, removal of projects that involve like-for-like replacements, and demonstration that qualifying program measures exceed code-based energy efficiency requirements associated with original construction or subsequent upgrades.

- Regarding eligibility, the evaluation team recommends that the PAs clearly document the energy efficiency action that is being performed and ensure that program rules are followed. Projects should have an identifiable and documented case for energy efficiency claims and application documentation should adequately explain how a given project saves energy.
- 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 22 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, 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.

**There were a number of cases where ISP or code-based baseline determination rendered a project ineligible. In these cases where project 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 internal communication and coordination efforts for disseminating, implementing and overseeing implementation of CPUC guidance should be increased.***
- **Recommendation:** ***To improve project eligibility screening it is recommended that the PAs ensure that incented measures exceed the ISP / code baseline. As such, it is important that the PAs spend adequate time documenting the appropriate project type and project baseline when establishing eligibility.*** The PAs are encouraged to examine Appendix F, which includes a list of every project where the evaluation overturned the PA specified project type or baseline type.

6.1.2 Project Operating Conditions

The operating conditions discrepancy factor accounted for 47 percent of all downward evaluation GRR result adjustments. While it is acknowledged that PAs cannot be aware of all changes in operating conditions that occur after incentives are paid, some aspects of operating conditions estimation can be addressed through improvement in program implementation activities and quality control.

Finding: Changed Operating Conditions for Projects

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.

The evaluation found that all PAs did not make adequate use of ex-ante data to inform operating conditions. The accuracy of PA savings estimates were deeply impacted by this discrepancy factor; for SCG operating conditions accounted for 73 percent of all downward adjustments to ex-ante claims.

- **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 improved 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, operational parameters and production levels, and ensure that data used to derive operating profiles is adequately representative of all operating conditions.

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 used in updating ex ante savings after equipment installation 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 (and other variables like weather where applicable) 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.
- 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 Project Baseline Specification

Improper baseline specification resulted in substantial adjustments to ex-ante savings claims for both electric and gas projects. These adjustments largely arose from a lack of conformance with CPUC baseline policy and guidance surrounding ISP, regressive baseline rules, full consideration of relevant codes, and a lack of documentation and data supporting the pre-existing conditions.

While all PAs had projects with deficiencies in baseline selection, baseline issues led to substantial downward savings adjustments to PG&E and SCE ex-ante saving estimates.

Finding: PA Baseline Changed by Evaluation

There was generally good agreement on project type and project baseline when comparing PA and evaluator selections (68 percent agreement across all PAs and projects). Add on, ER and ROB projects were the most commonly overturned project types across all PAs.

- **Recommendation**: ***Increase efforts to ensure 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 and documentation 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 remaining useful life (RUL) documentation in project application files should be a continued area of focus. 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. RUL is also needed to establish ROB and NR determination. For all early replacement (ER) projects, the PAs should provide and clearly document the RUL of the pre-existing equipment, in order to establish whether or not the removed system would fail. 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 ensure 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.

- **While the evaluation views documentation of RUL in the application paperwork to be critical, it is important to point out that population of RUL in the tracking system has led to confusion regarding claims and whether or not dual-baseline is implied in such cases. ***It is recommended that the utilities use better rules to populate RUL in the tracking system and/or agree on how best to interpret/communicate project type, working with the CPUC.***
- **SCG treatment of RUL in the tracking system provides a good example. RUL is only populated where the claim is clearly called out as ER.
- **All SCE tracking system records that were evaluated include an RUL, and RUL is set equal to EUL for all records that are not marked ER. This and the fact that other tracking system variables suggest a variety of project types, such as new construction and retrofit, seems to indicate that these entries are erroneous.
- **Also, for SDG&E a great many records include RUL, including records where other tracking system variables suggest a variety of project types.

- **Recommendation**: ***Clearly identify project event in terms of natural replacement, replace on burnout, early replacement, new construction, add-on equipment, and system optimization, 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.
- **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.

Finding: Greater PA Effort Needed for Proper Baseline Selection

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.

- ****Recommendation**: ***Where applicable, the PAs need to carefully investigate and document the age, condition and functionality of existing equipment and operations, and use these to establish proper baselines. Furthermore, when baseline conditions are defined by the pre-existing systems the PAs should utilize measured data to define those conditions where possible, and thoroughly document the pre-existing conditions for the purposes of establishing baseline.*** This is also relevant for ER claims. For ER claims preponderance of evidence should be used to accept or reject program induced early retirement. Existing equipment efficiency levels are needed to address regressive baseline policy.

6.1.4 Project 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.

Downward adjustments to ex-ante claims due to calculation methods were important for all PAs except SCG, where this factor accounted for just 4 percent of downward adjustments. Calculation method issues were most significant for PG&E and SDG&E.

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 23 percent of projects included in the evaluation gross impact sample. The evaluators often found it necessary to modify PA models (57 percent of projects) and / or inputs and assumptions (approximately 53 percent of projects). 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.
- ****Recommendation:** ***It is recommended that the PAs review all modeling weaknesses and areas for improvement noted in Section 3.5.***

Finding: PA Models Were Not Always Calibrated Using Observed Conditions

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 calibrate models and true-up savings based upon post-installation data, such as equipment usage profiles, equipment specifications and model inputs.*** The PAs should also make better use of available post-installation M&V data, including measured usage data and model inputs such as temperature settings and equipment operating schedules. Metering, EMS and SCADA data should be used to confirm or derive model inputs, such as operating condition, and to calibrate models.
- **Recommendation:** ***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 0.51. NTGR results indicate a medium[73] level of free ridership and a resulting medium level of program influence. Although this demonstrates a slight decline from the PY2013 NTGR of 0.54, this value continues to be similar in magnitude to NTGRs from the past several evaluation cycles, as shown in Table 6-3. 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.

[73] 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).

Table 6-3: Statewide Industrial Custom Program[74] Evaluation Net to Gross Ratios, Program Years 1998-2014

(1 – Free Ridership)	1998	1999	2000	2001	2002	2004 - 2005	PY2006-2008		2010-2012	2013	2014
							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	Statewide MMBtu - 0.51

*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. Further, for those projects already at an advanced stage, where equipment had already been budgeted, program influence was very low. There were also instances where incentives were provided to firms that were already very advanced in their adoptions of energy efficiency, such as companies with established energy efficiency procurement policies or mandates, including national chain and big box stores.

- **Recommendation:** ***Adopt procedures to identify and affect projects with low program influence.*** The 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.

[74] 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.

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 known free riders by upselling to higher efficiency levels, multi-measure solutions and continuous energy improvement.*** 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 automation of a manual process, 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 by setting incentive levels to maximize net (not gross) program impacts.*** *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,[75] 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

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

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 outreach channels to engage individual customers and encourage a long-term energy efficiency-based focus.*** Use a broad mix of program features and outreach 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.

6.3 Other Findings and Recommendations

Additional findings and recommendations are presented in this section that stem from general trends that were identified in this report. Subsections include the following:

- Overarching Considerations
- Project Type and Related Baseline Selection
- Project Cost-Effectiveness and Lifecycle Savings Estimation

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 an attempt to observe any effects of the EAR process. Given that the EAR process began in earnest in January 2012, the 2013+ period represents custom project applications under full EAR influence.

*****Finding: Importance of Documentation and M&V***

For three of the PAs, the PPA assessment found limited evidence of improvement in PA performance in the 2013+ period relative to pre-2013. SCG was an exception. SCG efforts to document and consistently in report in the 2013+ period were apparent. Highlights include the following for SCG:

- Successfully screened projects for eligibility
- The project type and related baseline designations were reasonably accurate
- EUL was well documented in the applications
- Documented and recorded incremental project costs in the 2013+ time period and showed improvement over the pre-2013 period
- It was also noted in this evaluation that SCG is successful in transferring documented conditions from the application forms and reports into the tracking system and does so in a consistent manner

PG&E also performed well in terms of documentation and tracking. For SCE in particular the appearance is that this is still a new process, and that relevant CPUC guidance on project treatment and documentation has not yet been disseminated adequately throughout their organization. Both SDG&E and SCE were found to populate RUL in the tracking system in a haphazard fashion, and attention and improvement is needed in this area.

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.

**The 2013 and 2014 PPA results, combined with GRR and NTGR findings, provide a solid baseline from which to continue tracking PA performance. Unfortunately for SCG the documentation, screening and accuracy considerations noted above did not translate into relatively good GRR results. In fact, SCG had the lowest MMBtu-based ex-ante LC and FY GRRs. For SCG the factor that drove down GRR results was misspecification of operating conditions. In many instances this resulted from a lack of M&V.

6.3.2 Project Type and Related Baseline Selection

Patterns emerged when examining project type and baseline agreement between PA and evaluator designations.

Finding: Add-on, 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 add-on, ER, and ROB project types. Similarly, the evaluation team often examined the efficiency level of replaced equipment which helped to identify instances of regressive baselines and subsequent project ineligibility. Finally, 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.

6.3.3 Project Cost-Effectiveness and Lifecycle Savings Estimation

*****Finding: Better ex-ante documentation is needed supporting cost-effectiveness and lifecycle savings estimation.***

Key variables, such as incremental cost, second period savings for ER projects, RUL and EUL, are often missing, incorrect and/or not well documented in the applications.