2015 CUSTOM IMPACT EVALUATION INDUSTRIAL, AGRICULTURAL, AND LARGE COMMERCIAL

Final Report

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

Prepared by:

Itron

1111 Broadway
Suite 1800
Oakland, CA 94607
www.itron.com/consulting

May 3, 2017

With Assistance from:
Energy and Resource Solutions
Energy Metrics
Michaels Energy
Katin Engineering Consulting
DNV GL
PWP, Inc.

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY 1-1

1.1 NEED FOR STUDY 1-1
1.2 CUSTOM PROGRAMS 1-1
1.3 APPROACH 1-2
1.4 RESULTS 1-4
1.5 RECOMMENDATIONS 1-8
1.6 CONTACT INFORMATION 1-9

2 TECHNICAL SUMMARY 2-1

2.1 CUSTOM IMPACT EVALUATION PORTFOLIO CONTEXT AND SAMPLE SIZES 2-2
2.2 HIGH-LEVEL CUSTOM GROSS IMPACT RESULTS 2-4
2.3 HIGH LEVEL CUSTOM NET-TO-GROSS RESULTS 2-8
2.4 NET EVALUATION REALIZATION RATE RESULTS 2-10
2.5 SUMMARY OF FINDINGS AND RECOMMENDATIONS 2-13

3 INTRODUCTION AND BACKGROUND 3-1

3.1 BACKGROUND 3-2
3.2 STUDY OBJECTIVES AND RESEARCHABLE ISSUES 3-8
3.3 STRUCTURE OF THE REPORT 3-10

4 GROSS IMPACT RESULTS 4-1

4.1 PROJECT-SPECIFIC GROSS IMPACT SUMMARY 4-2
4.2 PA GROSS REALIZATION RATE RESULTS 4-3
4.3 EAR OVERLAP SENSITIVITY ANALYSIS 4-12
4.4 DISCREPANCY ANALYSIS 4-15
4.4.1 Summary of Discrepancy Factor Impact 4-16
4.4.2 Discrepancy Factor Assessment for Projects with the Lowest GRRs 4-16
4.4.3 Assessment of Downward and Upward Adjustments to Gross Ex-Ante Savings by Discrepancy Factor 4-19
4.4.4 Categorical Explanation for Primary Discrepancy Factors 4-23
4.5 EVALUATION SUGGESTIONS AND CONSIDERATIONS TO ADDRESS THE MOST INFLUENTIAL DISCREPANCY FACTORS 4-28
4.5.1 Calculation Methods 4-29
4.5.2 Inappropriate Baseline 4-32
4.5.3 Ineligible Measures 4-36
4.5.4 Operating Conditions 4-37

5 NTG RESULTS 5-1

5.1 NUMBER OF COMPLETED SURVEYS 5-1
5.2 WEIGHTED NTG RESULTS 5-2
5.2.1 PG&E Combined Electric and Gas 5-3
5.2.2 SCE Electric 5-4
5.2.3 SDG&E Combined Electric and Gas 5-5

5.2.4 SCG Gas 5-6
5.2.5 Comparison of 2010-12, 2013, 2014, and 2015 NTG Results by PA — Combined MMBtu 5-7
5.3 NTG SENSITIVITY ANALYSIS 5-9
5.4 KEY FACTORS INFLUENCING NTGRS 5-9

6 PROJECT PRACTICES ASSESSMENTS 6-1

6.1 INTRODUCTION 6-1
6.2 OVERVIEW OF THE PROJECT PRACTICES ASSESSMENT 6-2
6.3 PROJECT PRACTICES ASSESSMENT RESULTS 6-3
6.3.1 Project Eligibility Assessments 6-3
6.3.2 Project Type Assessment 6-8
6.3.3 Project Baseline Assessment 6-15
6.3.4 Project Baseline Ratings 6-21
6.3.5 EUL Assessment 6-26
6.3.6 Calculations Assessment 6-29
6.3.7 Inputs and Assumptions Assessment 6-32
6.3.8 Incremental Cost Assessment 6-36

7 DETAILED EVALUATION FINDINGS AND RECOMMENDATIONS 7-1

7.1 GROSS IMPACT-RELATED FINDINGS AND RECOMMENDATIONS 7-4
7.1.1 Underperforming Projects 7-4
7.1.2 Project Calculation Methods 7-6
7.1.3 Project Baseline Specification 7-9
7.1.4 Project Operating Conditions 7-12
7.1.5 The State of Ex-Ante M&V 7-14
7.2 NET-TO-GROSS / PROGRAM INFLUENCE ISSUES 7-15

LIST OF FIGURES

Figure 1-1: Custom Impact Evaluation Share of Statewide 2015 First Year Gross Energy Claims 1-2
Figure 1-2: Mean Lifecycle Gross Realization Rates by IOU for electric and gas energy savings and electric demand 1-5
Figure 1-3: Weighted Net-to-Gross Ratios by IOU 1-6
Figure 2-1: Custom Impact Evaluation Share of Statewide PY2015 First Year gross Energy Efficiency claims* 2-3
Figure 2-2: Mean Lifecycle Gross Realization Rates by PA and Energy Metric (MMBtu and kW) 2-5
Figure 2-3: Summary of Discrepancy Factors Resulting in Downward and Upward Adjustments to First Year Ex-Ante MMBtu Impacts - All PAs 2-7
Figure 2-4: Weighted Net-to-Gross Ratios by PA 2-8
Figure 3-1: 2015 IALC Roadmap First year adjusted gross ex-ante Savings Relative to Portfolio savings 3-5
Figure 3-2: 2015 First Year adjusted gross ex-ante Savings Assigned to the IALC Roadmap 3-7
Figure 4-1: Lifecycle Ex-Ante and Ex-Post Combined Electric and Gas gross Savings (MMBtu) for Sampled Projects 4-3
Figure 4-2: Lifecycle Gross Realization Rate Results by PA for Combined Electric and Gas Savings (MMBtu) and for Peak Electric Demand (kW) 4-4

Figure 4-3: Comparison of 2010-12, 2013, 2014, and 2015 Weighted MMBtu LC GRR Results 4-8
Figure 4-4: Comparison of 2010-12, 2013, 2014, and 2015 Weighted kW LC GRR Results 4-9
Figure 4-5: Ex-post Upward and Downward Adjustments to First year Gross Ex-ante MMBtu for Sampled Projects - All PAs 4-22
Figure 4-6: Distribution of Downward and Upward Adjustments to First year gross Ex-ante MMBtu by Discrepancy Factor - All PAs 4-23
Figure 4-7: Most Influential Discrepancy Factors that Caused Downward Adjustments for PG&E 4-24
Figure 4-8: Most Influential Discrepancy Factors that Caused Downward Adjustments for SCE 4-25
Figure 4-9: Most Influential Discrepancy Factors that Caused Downward Adjustments for SDG&E 4-26
Figure 4-10: Most Influential Discrepancy Factors that Caused Downward Adjustments for SCG 4-27
Figure 5-1: Comparison of 2010-12, 2013, 2014, and 2015 Weighted MMBtu* NTGR Results 5-8
Figure 5-2: Program Influence by Market Segment for Low and High NTGR Categories (2013-15) 5-13
Figure 6-1: Comparison of Project Eligibility Across Program Years 6-7
Figure 6-2: PG&E Count of Project Types – Ex-Ante and Ex-Post 6-12
Figure 6-3: SCE Count of Project Types – Ex-Ante and Ex-Post 6-13
Figure 6-4: SDG&E Count of Project Types – Ex-Ante and Ex-Post 6-14
Figure 6-5: SCG Count of Project Types – Ex-Ante and Ex-Post 6-15
Figure 6-6: PG&E Distribution of Project Baselines – Ex-Ante and Ex-Post 6-18
Figure 6-7: SCE Distribution of Project Baselines – Ex-Ante and Ex-Post 6-19
Figure 6-8: SDG&E Distribution of Project Baselines – Ex-Ante and Ex-Post 6-20
Figure 6-9: SCG Distribution of Project Baselines – Ex-Ante and Ex-Post 6-21
Figure 6-10: Mean Project Baseline Documentation Quality Ratings by PA 6-23
Figure 6-11: Comparison Mean Project Baseline Ratings by PA Across Program Years 6-25
Figure 6-12: Calculations Methods Assessment by PA 6-30
Figure 6-13: Comparison Mean Calculation Method Ratings by PA Across Program Years 6-31
Figure 6-14: Inputs and Assumptions Assessment by PA 6-34
Figure 6-15: Comparison Mean Input and ASsumption Ratings by PA Across Program Years 6-35

LIST OF TABLES

Table 1-1: Summary of Custom Evaluation Sample Sizes and MMBTu Sample Percent Representation by PA 1-4
Table 1-2: Ex Ante and Ex Post Net Lifecycle Mwh Savings and Realization Rates 1-7
Table 1-3: Ex Ante and Ex Post Net Lifecycle Mw Savings and Realization Rates 1-7
Table 1-4: Ex Ante and Ex Post Net Lifecycle therm Savings and Realization Rates 1-7
Table 2-1: Summary of Custom Evaluation Sample Sizes by PA 2-4
Table 2-2: Mean Lifecycle Gross Realization Rates by PA and Energy Metric (MMBtu and kW) 2-6
Table 2-3: Weighted Net-to-Gross Ratios by PA 2-9
Table 2-4: PG&E Lifecycle Net Realization Rate Estimates and Comparisons 2-11
Table 2-5: SCE Lifecycle Net Realization Rate Estimates and Comparisons 2-12
Table 2-6: SDG&E Net Realization Rate Estimates and Comparisons 2-12
Table 2-7: SCG Lifecycle Net Realization Rate Estimates and Comparisons 2-13

Table 2-8: Summary of Key Recommendations 2-14
Table 3-1: 2015 First year adjusted gross ex-Ante Energy Impacts by PA for Projects in the IALC Roadmap, and in the Portfolio 3-4
Table 3-2: 2015 First Year ajdusted gross ex-ante Energy Impacts for the IALC Roadmap, by Group and PA 3-6
Table 3-3: 2015 Unadjusted First Year Gross Savings Estimates Evaluated by IALC Custom, by PA 3-8
Table 3-4: Overall Organizational Structure of Report 3-10
Table 4-1: Custom Evaluation Gross M&V Sample Disposition by PA 4-2
Table 4-2: Weighted Project Lifecycle and First Year Gross Realization Rates by PA and Energy Metric (MMBtu and kW) 4-5
Table 4-3: 2010-2012, 2013, 2014, and 2015 Weighted Project Realization Rates by PA and Energy Metric (MMBtu and kW) 4-7
Table 4-4: Project Lifecycle Realization Rates by Strata and Sample Domain 4-11
Table 4-5: MMBtu Realization Rates for Sampled EAR Projects 4-14
Table 4-6: Discrepancy Factors for Projects with Zero or Negative gross Ex-Post MMBtu savings 4-18
Table 4-7: Records with Ex-Post Downward and Upward Adjustments to First year gross Ex-Ante MMBtu Impacts, by Discrepancy Factor; Statewide and by PA 4-20
Table 5-1: Completed Surveys by Program Administrator 5-2
Table 5-2: Surveyed Projects Removed from NTG Analysis 5-3
Table 5-3: Weighted Net-to-Gross Ratios for PG&E – Combined Electric and Gas 5-3
Table 5-4: Weighted Net-to-Gross Ratios for SCE – Electric 5-4
Table 5-5: Weighted Net-to-Gross Ratios for SDG&E – Combined Electric and Gas 5-5
Table 5-6: Weighted Net-to-Gross Ratios for SCG – Gas 5-6
Table 5-7: Comparison of 2010-12, 2013, 2014, and 2015 Weighted MMBtu* NTGR Results 5-8
Table 5-8: Results of NTG Sensitivity Analysis – All PAs and All Sample Points 5-9
Table 5-9: Key Factors Affecting NTGRs for all PAs 5-12
Table 6-1: Summary of Ineligible Measures, and Ex-Post M&V Conclusions Why Measures Are Ineligible 6-5
Table 6-2: Frequency of Ex-Ante and Ex-Post Agreement on Project Type by PA and Program Year 6-10
Table 6-3: PA vs. Evaluation Specified Project Type – All PAs 6-11
Table 6-4: Frequency of Ex-Ante and Ex-Post Agreement on Project Baseline by PA and Program Year 6-16
Table 6-5: PA vs. Evaluation Specified Project Baseline – All PAs 6-17
Table 6-6: Project Baseline Documentation Quality by PA 6-22
Table 6-7: EUL Assessment by PA 6-27
Table 6-8: Calculations Methods Assessment by PA 6-29
Table 6-9: Inputs and Assumptions Assessment by PA 6-33
Table 6-10: Incremental Cost Documentation by PA 6-36
Table 7-1: 2010-2012, 2013, 2014 and 2015 Weighted Project Lifecycle Realization Rates by PA and Energy Metric (MMBtu and kW) 7-2
Table 7-2: Comparison of 2010-12, and 2013, 2014 and 2015 Weighted MMBtu* NTGR Results 7-3
Table 7-3: Statewide Industrial and Custom Program Evaluation Net to Gross Ratios, Program Years 1998-2015 7-15

1 EXECUTIVE SUMMARY

1.1 NEED FOR STUDY

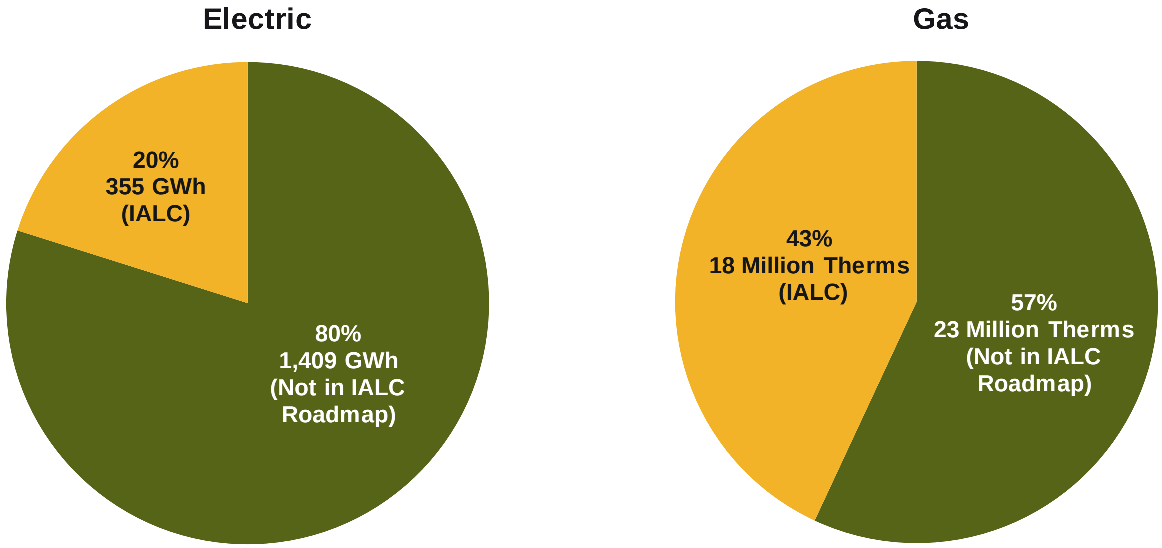
For decades, the California investor owned utilities (IOUs)[1] have offered energy efficiency programs. These programs attempt to influence customers to install energy efficient equipment and systems (such as compressed air equipment used in an assembly plant) by providing information, rebates, and other forms of monetary incentives. The term “custom programs” refers to energy efficiency programs that typically involve complex equipment and systems where savings are calculated by the IOUs individually for each project. In this study, custom energy efficiency projects that received monetary incentives in 2015 from the IOU programs are evaluated. A combination of engineering, social science and statistical analysis is used in order to develop independent savings estimates and report on lessons learned for the programs.

1.2 CUSTOM PROGRAMS

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. As shown in Figure 1-1, energy savings claims associated with the scope of this custom evaluation represent a significant contribution to the overall savings claims for the IOUs’ energy efficiency programs, accounting for about 20 percent of statewide electric savings claims and 43 percent of statewide gas savings claims during 2015. Given that the IOUs spend more than $800 million per year across all programs, and that custom projects are such an important component of the savings claimed, it is crucial that custom programs and projects be evaluated through recurring cycles of examination, followed by subsequent steps to improve program performance and accomplishments.

[1] Pacific Gas and Electric (PG&E), Southern California Edison (SCE), San Diego Gas and Electric (SDG&E) and Southern California Gas (SCG)

FIGURE 1-1: CUSTOM IMPACT EVALUATION SHARE OF STATEWIDE 2015 FIRST YEAR GROSS ENERGY CLAIMS



1.3 APPROACH

The goals and objectives of this Custom evaluation are: to verify and validate the energy efficiency savings claims reported from IOU energy efficiency programs; to provide feedback on how well program procedures and savings calculation methods align with the CPUC's energy efficiency policies, requirements, and expectations; and to provide recommendations on how custom programs can be improved or refined.

To support these goals and objectives, the Custom evaluation sampled projects and conducted site-level evaluation -- 148 gross impact points and 208 net impact points, where a point or "sampling unit" is defined as an individual project installed at a specific participating customer site. Gross impact estimates for sampled projects were based on field inspections, measurements, and extensive engineering analysis (i.e., measurement and verification, or M&V). These M&V evaluation activities yield verified, independent estimates of savings for each project in the sample. The "net-to-gross" (NTG) evaluation consisted of interview-based evaluation for a sample of participating customers, with a goal of quantifying the influence of the program in garnering energy savings.

The evaluation compares the initial energy savings claim made by the programs to the evaluation's results, which are developed using independent calculations. The initial savings are often referred to as ex-ante savings, because these are the savings values before (ex-ante) the evaluation is conducted. The evaluation savings values are then referred to as the ex-post savings, because these are the savings values developed after (ex-post) the evaluation was able to assess the energy saving measures installed by the programs.

The ratio of the ex-post (evaluation estimated) to ex-ante (program estimated) savings is referred to as the "realization rate," or the rate at which ex-ante savings are realized through the evaluation. From the representative gross impact sample, the evaluation can determine an average realization rate for each IOU, using a statistical method known as ratio estimation.

Through the NTG activities, the evaluation also examines how successful the IOU programs were in influencing utility customers to install energy efficient measures that would not have been installed if the programs had not existed. Customers that would have installed the same energy efficient equipment (and at the same time) in the absence of the program are considered free riders. They are referred to as free riders because they are receiving incentives from the programs for actions they would have undertaken without the program's existence. The evaluation examines both the "gross" amount of savings derived among all participants, and the savings that is generated "net" of free riders.

This evaluation-developed estimate of the ratio between the net and gross levels of savings is known as the net-to-gross ratio (or NTGR). To estimate the NTGR, a representative sample of participants are telephone surveyed and asked several questions regarding the program's influence on their decision to install the energy efficient equipment. The survey examines various factors related to the program and other non-program factors.[2] The survey also examines what the customer would likely have done in the absence of the program.

These survey question responses determine how likely it is that the program has influenced the customer's decision to install program qualifying high efficiency equipment or systems, and conversely, how likely it is that the participant was a free rider. For the sample of telephone surveyed participants, the NTGR is first estimated, then expanded using ratio estimation to yield a mean value by IOU, and subsequently multiplied by the programs overall gross savings value for each IOU to estimate the programs overall net savings value.

The ultimate goal of this evaluation is to arrive at an estimate of the ex-post net lifecycle energy and demand savings. The ex-post net lifecycle energy savings represents the gross savings that accumulate over the life of all of the custom projects installed in 2015, minus (net) the free riders. The definition of ex-post net demand savings is similar to the energy savings definition but represents the savings achieved during periods of high electricity use that ordinarily occur during high temperature summer "peak" events, which typically only persist for several hours in a given day, but may occur on more than one day. Program-based estimates of ex-ante net lifecycle energy and demand savings are also presented in the evaluation, providing an opportunity to directly compare ex-post and ex-ante savings.

[2] NTG, as reported here, is inclusive only of free ridership effects (1-FR) and does not include spillover or market effects.

The evaluation set specific sampling targets for each IOU, for both M&V and telephone surveys. The original sample design was to have 200 NTG project surveys completed, 150 of which were to overlap with the 150-point gross M&V sample. For M&V the PG&E and SCE targets were roughly 40 points each, and those for SDG&E and SCG were roughly 30 points each; and for net 55 and 45 points each, respectively. Table 1-1 presents the resulting sample sizes, as well as the percentage of total custom project energy savings claims (gas and electric) that are represented by each sample.

TABLE 1-1: SUMMARY OF CUSTOM EVALUATION SAMPLE SIZES AND MMBTU SAMPLE PERCENT REPRESENTATION BY PA

PA	Completed M&V Points (n) and MMBtu Sample Percent Representation		Completed NTG Points (n) and MMBtu Sample Percent Representation	
PG&E	42	28%	54	29%
SCE	43	47%	54	42%
SDG&E	33	65%	44	64%
SCG	30	72%	56	66%
All PAs	**148**	**41%**	**208**	**39%**

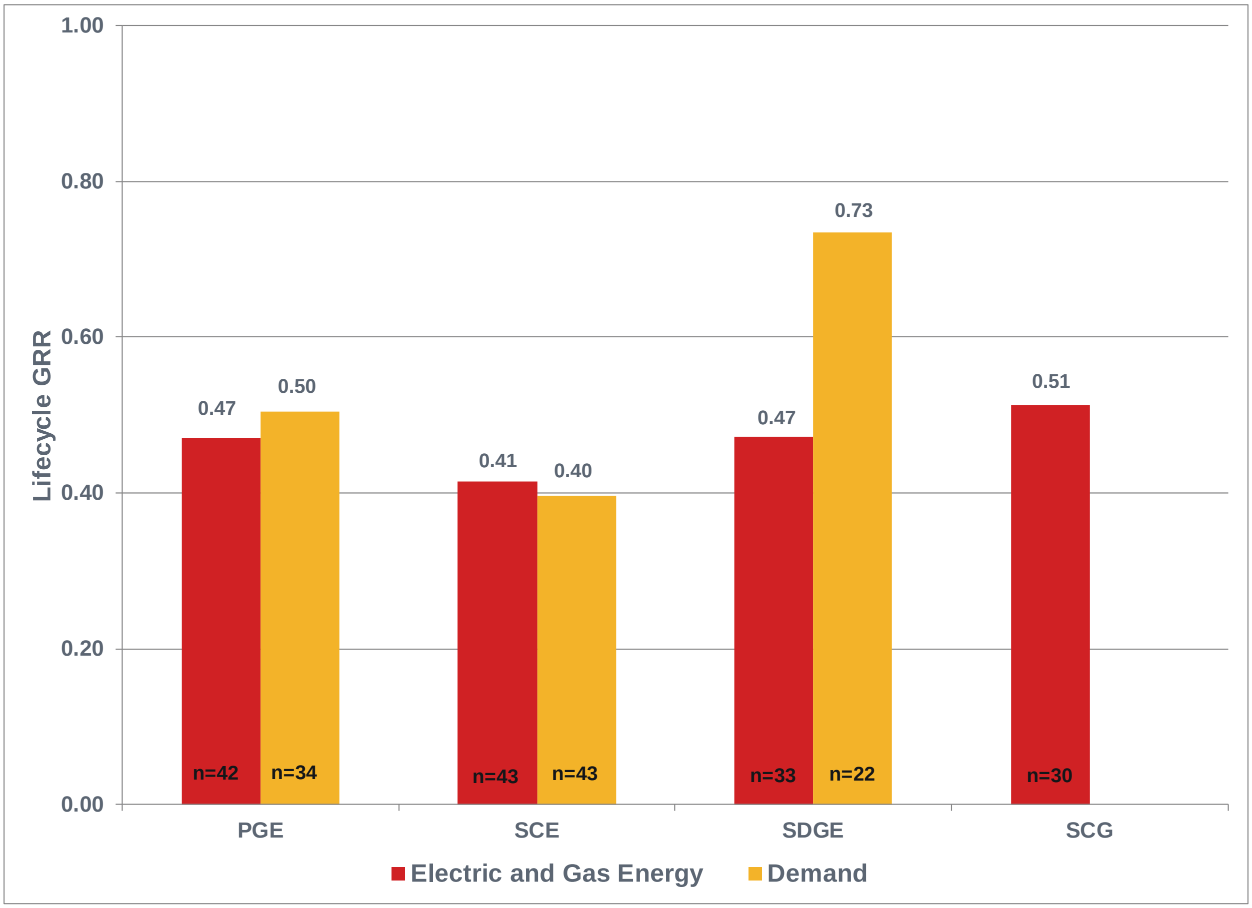
1.4 RESULTS

The results of this evaluation are summarized below in three different ways for each IOU. First, the evaluation gross impact results are summarized based on a presentation of realization rates, as defined above in Section 1.3. Second, the evaluation net impact results are summarized based on a presentation of NTGRs, which were also defined above in Section 1.3. Finally, the results of this evaluation are presented using a comparison between the ex-post (evaluation) and ex-ante (claimed) net lifecycle savings estimates, and the ratio of those two values, which results in a net realization rate. Figure 1-2 and Figure 1-3, and Tables 1-2, 1-3 and 1-4 below present these results.

The mean lifecycle realization rates by IOU are presented in Figure 1-2 for both energy and demand savings, and are less than 0.52 for all but one result (the realization rate for SDG&E demand is 0.73). These results indicate that there is substantial room for improvements to ex-ante calculated savings estimates.

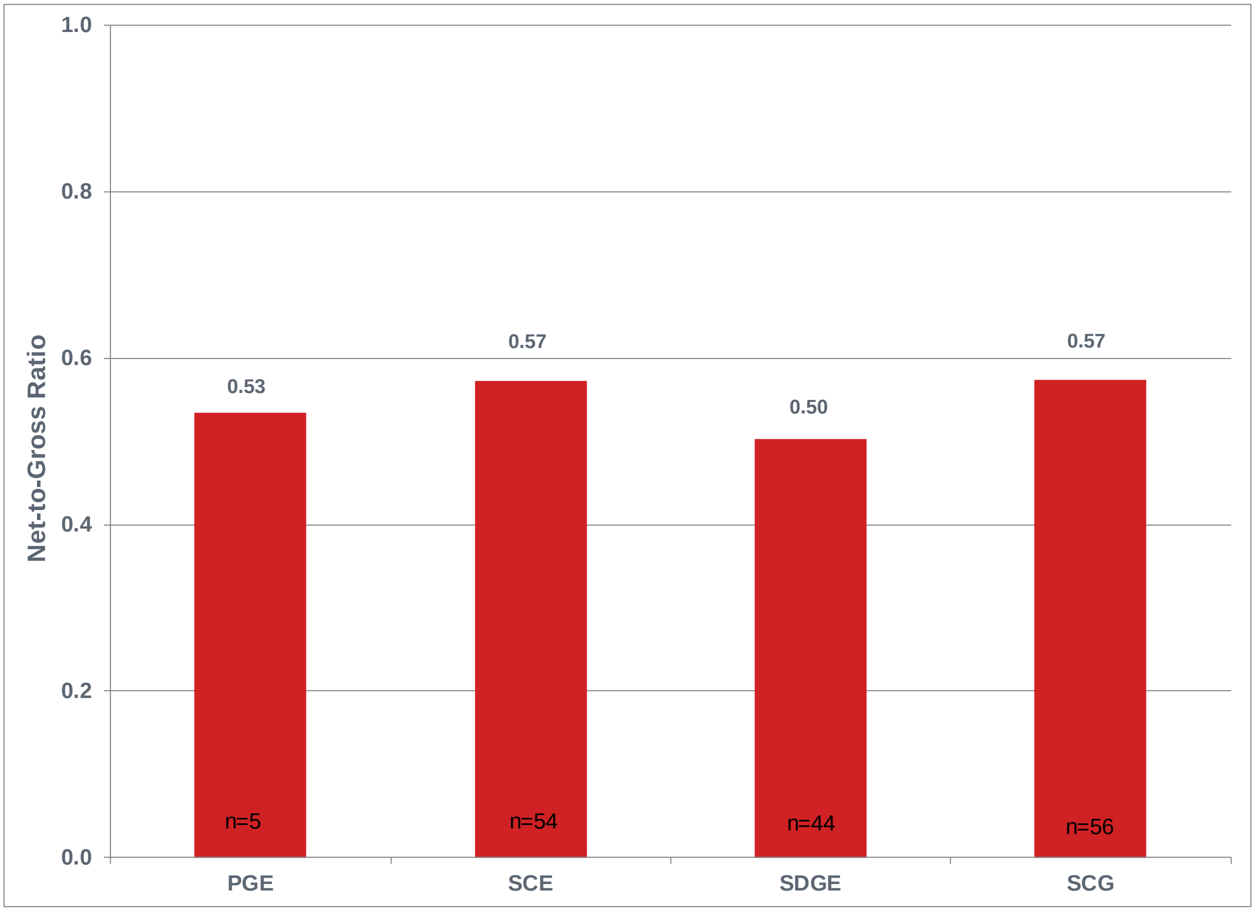
These results indicate that the achieved ex-post savings are much lower than ex-ante claimed savings, and this evaluation examines the discrepancy factors that lead to a reduced ex-post savings. The four principal reasons that ex-ante gross impacts differ from ex-post results are: (1) the IOUs' calculation methods, (2) baseline specification, (3) ineligible measures, and (4) use of observed operating conditions. A realization rate result of 1.0 would be indicative of equivalent ex-ante and ex-post project treatment and approach, and would yield savings that are roughly double what was achieved in 2015. This represents a difficult, but not impossible milestone to achieve. Efforts are needed to close this gap. Recommendations to do so are presented in Section 1.5 below.

FIGURE 1-2: MEAN LIFECYCLE GROSS REALIZATION RATES BY IOU FOR ELECTRIC AND GAS ENERGY SAVINGS AND ELECTRIC DEMAND



NTGR results by IOU are presented in Figure 1-3. As noted above, NTGRs are indicative of the influence of the program in driving customers to install energy efficiency projects, given that there are a host of other factors that can also influence project design and equipment specifications, including business needs, market conditions and environmental compliance with regulations, among others. An NTGR result of 1.0 would indicate that the program alone influenced program participants to install projects that were funded by IOU programs. The NTGR results below (0.50 to 0.57) indicate only a moderate level of program influence and substantial room for improvement. This evaluation provides recommendations and feedback surrounding program designs that should yield higher NTGR results.

FIGURE 1-3: WEIGHTED NET-TO-GROSS RATIOS BY IOU



Electric energy net savings results and net realization rates are presented by IOU in Table 1-2. Again, net savings realization rates are simply a ratio of ex-post (evaluated) to ex-ante (claimed) net savings, providing a measure of program success in achieving claimed electric energy savings, which we see hovers around 40 percent. The ex-post estimates of net savings incorporate both the gross realization rate and NTGR results from Figure 1-2 and Figure 1-3, as multipliers applied to ex-ante savings estimates. The ex-ante net savings claims also incorporate gross savings realization rate and NTGR estimates, but the product of the two terms has a smaller downward effect on claims compared to ex-post savings adjustments.

TABLE 1-2: EX ANTE AND EX POST NET LIFECYCLE MWH SAVINGS AND REALIZATION RATES

IOU	Electric Energy Lifecycle Net MWh Savings		
	Ex Ante (Claimed)	Ex Post (Evaluated)	Net Realization Rate (Ex Post/Ex Ante)
PGE	1,563,964	608,025	39%
SCE	1,119,486	428,716	38%
SDGE	145,170	62,761	43%
SCG	NA	NA	NA

Electric demand net savings results and net realization rates are presented by IOU in Table 1-3. While PG&E and SCE electric demand net realization rates are around 40 percent, SDG&E results approach 70 percent.

TABLE 1-3: EX ANTE AND EX POST NET LIFECYCLE MW SAVINGS AND REALIZATION RATES

IOU	Electric Demand Lifecycle Net MW Savings		
	Ex Ante (Claimed)	Ex Post (Evaluated)	Net Realization Rate (Ex Post/Ex Ante)
PGE	222	92	41%
SCE	175	66	37%
SDGE	21	14	68%
SCG	NA	NA	NA

Gas energy net savings results and net realization rates are presented by IOU in Table 1-4. Gas energy net realization rates range from 40 percent to 65 percent, thereby demonstrating slightly better performance in comparison with electric energy net savings estimates.

TABLE 1-4: EX ANTE AND EX POST NET LIFECYCLE THERM SAVINGS AND REALIZATION RATES

IOU	Gas Energy Lifecycle Net Therm Savings		
	Ex Ante (Claimed)	Ex Post (Evaluated)	Net Realization Rate (Ex Post/Ex Ante)
PGE	85,174,196	33,862,993	40%
SCE	222,343	128,375	58%
SDGE	8,605,371	4,236,528	49%
SCG	41,519,419	26,907,634	65%

1.5 RECOMMENDATIONS

Chapter 7 of the report provides conclusions and recommendations based on the ex-post evaluation of custom programs. The recommendations focus on suggestions for making improvements to ex-ante savings calculations, and thereby bring ex-post and ex-ante gross impact estimates into closer alignment. This can be achieved through targeted improvement to program processes, procedures and protocols that are designed to address the key discrepancy factors that lead to ex-ante and ex-post impact estimation differences. These recommendations are supported by specific findings which are featured in Chapters 4 and 5 of the report. These finding are summarized with each recommendation provided below.

- The IOUs should improve documentation and reporting of project effective useful life (EUL). The EUL defines the number of expected years of savings for a given project and associated equipment. It was found that the ex-ante estimates of project equipment life were greater on average than ex-post estimates, leading to exaggerated ex-ante lifecycle savings estimates. It is therefore recommended that the IOUs carefully review evaluation EUL conclusions/rationale in an effort to improve EUL claims and lifecycle savings estimates.
- The IOUs should emphasize improvements to address underperforming projects. Out of 148 M&V sample points, 30 projects, or 20 percent of the sample, were determined by the evaluation to either not save energy or, in some cases, to even increase energy use. The discrepancy factors that led to these poor results are identified in Chapter 4. Twenty-two of the 30 cases resulting in zero or negative savings were due principally to one of two factors – inappropriate baseline or ineligible measures. The IOUs need to improve program eligibility requirements and project baseline determination procedures in order to screen out projects that don't save energy – by updating manuals, providing training, and enhancing quality control procedures.
- The IOUs should also address needed improvements to calculation methods and protocols in an effort to enhance savings estimation accuracy. The ex-ante calculations for an array of projects were lacking in terms of the calculation method applied and incorporation of correct inputs that describe typical or representative operating conditions. The IOUs should review and improve impact methods and models and ensure adherence with savings estimation policies, guidelines and best practices. Furthermore, the IOUs should calibrate models and true-up savings based upon post-installation data, such as equipment usage profiles, equipment specifications, production records and model inputs.
- To reduce continued moderate free ridership, PAs should consider changes to program implementation procedures and features designed to increase program influence. These include: adopting procedures to identify and affect projects with low program influence; adjusting the set of technologies eligible for incentives; and implementing procedures to limit known free riders by upselling to higher efficiency levels, multi-measure solutions and continuous energy improvement.

- The focus of these efforts might best be directed to projects demonstrating the lowest NTGRs (0.36 or lower, and associated with the highest free ridership levels).
- In Chapter 5 it was shown that some customer segments are more likely to have projects with low NTGRs, such as water/wastewater facilities.

1.6 CONTACT INFORMATION

The ED Project Manager for this study was Ms. Katherine (Kay) Hardy. Itron served as the Prime Contractor managing this study, led by Mr. Kris Bradley.

The following is Ms. Hardy's and Mr. Bradley's contact information.

Firm	Lead	Contact Info
CPUC 505 Van Ness Ave San Francisco, CA 94102	Kay Hardy Energy Division DSM Evaluation Section	Phone: (415) 703-2322 Email: Katherine.hardy@cpuc.ca.gov
Itron, Inc 1111 Broadway, Suite 1800 Oakland, CA 94607	Kris Bradley, Director Consulting & Analysis	Phone: (510) 844-2818 Email: kris.bradley@itron.com

2 TECHNICAL SUMMARY

This report presents findings from the impact evaluation of the program year (PY) 2015 California Program Administrator (PA)[3] led energy efficiency programs, focusing on nonresidential custom measures.[4] This custom project impact evaluation is one of multiple California Public Utilities Commission (CPUC) evaluations of the PAs' 2015 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.[5] 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),[6] and a Project Practices Assessments (PPA) activity[7] 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, frozen version of the data necessary for the 2015 IALC evaluation was received by January 6th, 2017. The majority of the evaluation field work, data collection and sample point-level analysis activities took place in the second half of 2016, and the aggregate analysis and reporting was completed in the first half of 2017. This annual evaluation schedule results in the quickest feasible feedback to the PAs with regard to their program activities and supports the Efficiency Savings and Performance Incentive (ESPI) award.[8]

---

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

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

5 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.

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

7 Project Practices Assessment reviews were conducted for all completed measurement and verification (M&V) sample 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.

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

Three main evaluation activities support the findings and recommendations in this report: (1) M&V activities for estimating gross impacts for 148 projects, (2) telephone survey data collection supporting net to gross (NTG) estimation for a total of 208 projects, and (3) a total of 148 engineering reviews supporting PPA results.

2.1 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.[9] 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 and PY2015 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).

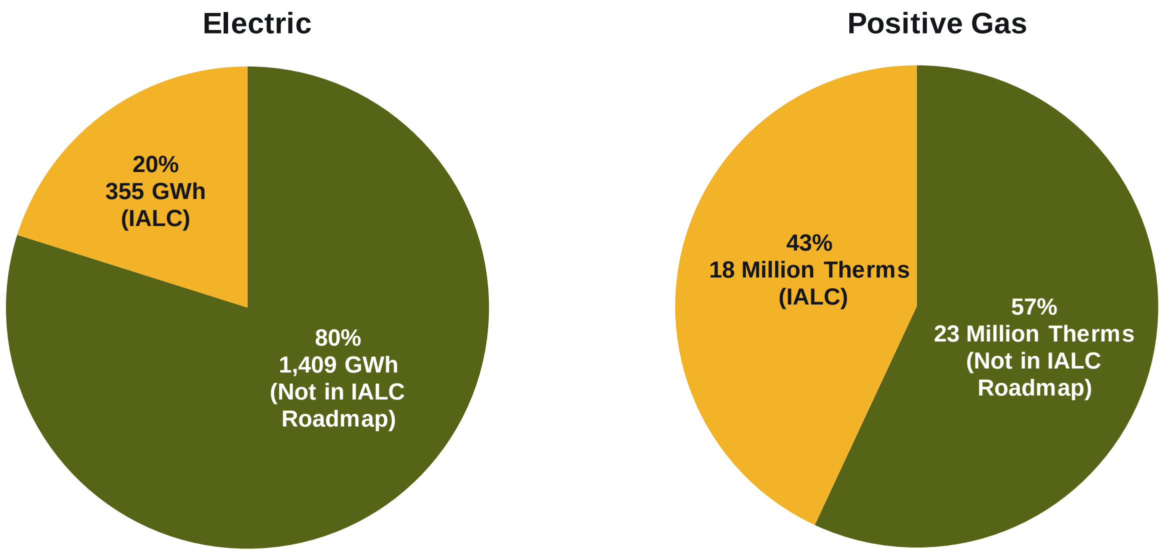
As shown in Figure 2-1, first year gross energy savings claims associated with the scope of this evaluation represent a significant contribution to the overall savings claims for the PAs' energy efficiency programs, accounting for about 20 percent of statewide electric savings claims and 43 percent of statewide gas savings claims[10] during PY2015. During this period, the PA tracking data for measures associated with this custom impact evaluation included thousands of records statewide with annual electric gross savings claims by the PAs totaling 355 GWh and annual gas gross savings claims totaling 18 million Therms.[11]

[9] Custom lighting measures are addressed in a separate impact study on nonresidential lighting, under the CPUC 2013-2015 Commercial Roadmap. Pump test claims were evaluated in the past and were not in scope for 2015.

[10] Excluding negative savings impacts associated with HVAC interactive effects.

[11] Of the 355 GWh and 18 million Therms associated with the IALC Roadmap, 10 percent of electric claims and two percent of gas claims were attributed to NRNC whole building projects. This equates to approximately eight percent of IALC savings on a combined MMBtu basis.

FIGURE 2-1: CUSTOM IMPACT EVALUATION SHARE OF STATEWIDE PY2015 FIRST YEAR GROSS ENERGY EFFICIENCY CLAIMS*



* "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,[12] where applicable for that PA. This approach resulted in the following four sampling domains for which gross realization rates[13] and net to gross (NTG) ratios[14] 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 148 gross impact points (consisting of 188 individual measures) and 208 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

---

12 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.

13 "Gross realization rate" is the evaluation gross savings estimate divided by the PA savings estimate.

14 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.

M&V sample. However, given customer willingness to participate and other factors, the final gross and net samples did not fully align. In total, 115 of the completed NTG sample points overlapped with the 148 evaluated gross M&V points. The total sample size (including main and backup points achieved) and sample percent representation of ex-ante MMBtu claims by PA are shown in Table 2-1 below. The sample percent representation for the gross M&V sample is based on lifecycle savings (LC), and the NTG sample percent representation is based on first year savings (FY).[15]

TABLE 2-1: SUMMARY OF CUSTOM EVALUATION SAMPLE SIZES BY PA

PA	Completed M&V Points (n)		Completed NTG Points (n)		Sample Percent Representation of Ex-Ante MMBtu Claims	
	Main	Backup	Main	Backup	M&V Sample (LC)	NTG Sample (FY)
PG&E	42	0	39	15	28%	29%
SCE	42	1	40	14	47%	42%
SDG&E	32	1	31	13	65%	64%
SCG	30	0	27	29	72%	66%
All PAs	**146**	**2**	**137**	**71**	**41%**	**39%**

2.2 HIGH-LEVEL CUSTOM GROSS IMPACT RESULTS

Figure 2-2 and Table 2-2 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 engineering estimate of gross savings divided by the PAs' ex-ante gross engineering 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 2-2 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 interval.

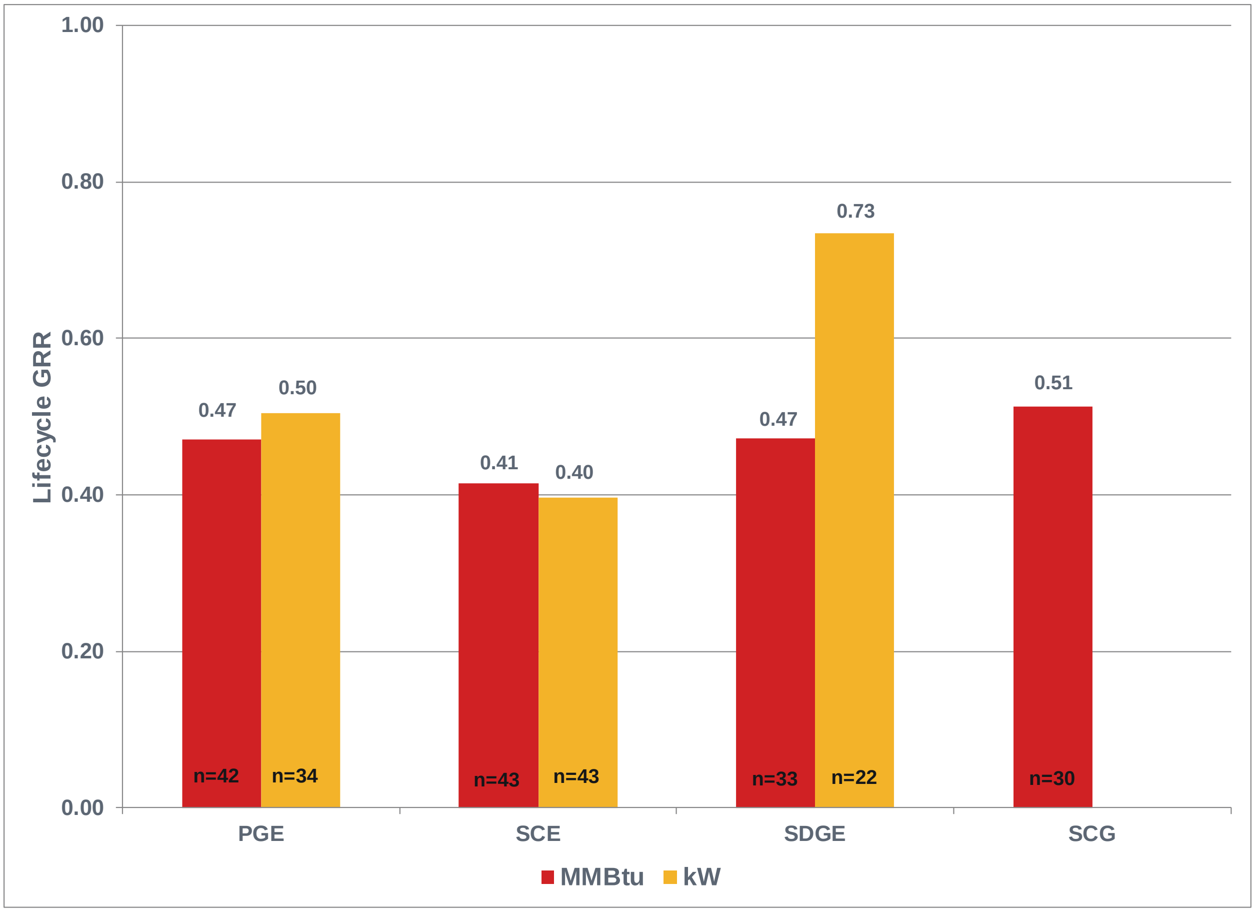
The mean lifecycle realization rates by PA and energy metric are less than 0.52 for all but one energy metric (the GRR for SDG&E kW is 0.73) and are similar, but generally lower than those from the 2010-2012, 2013, and 2014 evaluations (see weighted MMBtu comparison in Chapter 4). For all PAs, weighted lifecycle realization rates are lower than the corresponding first year realization rate.[16] Generally, evaluation lifecycle realization rates remain significantly below the 0.9 default ex-ante GRR adjustment

[15] The gross M&V sample was weighted by LC savings sample because it captures savings over the life of the measure. The NTG sample was weighted by FY savings because the results are based on decisions made by the customer prior to measure installation. Customers might be swayed by bill savings (operational savings), so FY is more appropriate than LC.

[16] Lifecycle gross realization rate results are lowered relative to first year results by differences between evaluation and ex-ante EUL determinations.

for custom programs. A significant number of projects (30 out of 148) were estimated to have negative and/or zero GRRs.

FIGURE 2-2: MEAN LIFECYCLE GROSS REALIZATION RATES BY PA AND ENERGY METRIC (MMBTU AND KW)



The error ratios for most domains (Table 2-2) are similar to the error ratios obtained in the *2010-2012 WO033 Custom Impact Evaluation Final Report* and the *2013 and 2014 IALC Custom Impact Evaluation Final Reports.*[17] In the 2015 sample, the relatively high standard deviation is largely 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 negative values 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 SCE's kW GRR consisted of 43 projects, two of which had LC GRRs of greater

[17] 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
http://www.calmac.org/publications/IALC_2014_Final_Report_April_2016.pdf

than 1.0 and 12 projects with LC GRRs of zero or less, resulting in an error ratio of 1.32.[18] While the precision of the 2015 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 2-2: MEAN LIFECYCLE GROSS REALIZATION RATES BY PA AND ENERGY METRIC (MMBTU AND KW)

Energy Metric	Population Count	Sample Count	% of LC Savings Sampled	LC Mean GRR	Error Ratio**	90% Confidence Interval	FY Mean GRR
				PG&E			
MMBtu*	1,351	42	28%	0.47	0.96	0.36 to 0.58	0.54
kW	1,066	34	20%	0.50	1.17	0.34 to 0.67	0.64
				SCE			
MMBtu*	765	43	47%	0.41	0.73	0.34 to 0.49	0.55
kW	702	43	38%	0.40	1.32	0.27 to 0.52	0.50
				SDG&E			
MMBtu*	98	33	65%	0.47	0.66	0.4 to 0.55	0.51
kW	67	22	49%	0.73	1.22	0.48 to 0.99	0.77
				SCG			
MMBtu	196	30	72%	0.51	0.92	0.38 to 0.64	0.53

* The primary sample was designed and selected based on ex-ante MMBtu savings estimates. The kW sample sizes are sometimes lower due to the fact that kW impacts were not always claimed by PAs for every project in the sample.

** A measure of the statistical variation in the gross realization rates. Note that this error metric only captures sampling error, not measurement error.

The four principal reasons that ex-ante gross impacts differ from ex-post results are: (1) the PAs' calculation methods, (2) baseline specification, (3) ineligible measures, and (4) use of observed operating conditions. These discrepancy factors were examined for all projects where they caused upward or downward adjustments to the ex-ante savings.[19] Of the 188 records (measures) studied, these discrepancy factors explain a portion of the downward adjustments in ex-ante savings for 31 percent, 18 percent, 11 percent and 33 percent of records, respectively.[20] For all downward adjustments, across all PAs, these four factors reduced the sample-wide ex-ante MMBtu savings estimates by a combined amount of 47 percent; 14 percent (calculation methods), 12 percent (inappropriate baseline), 12 percent

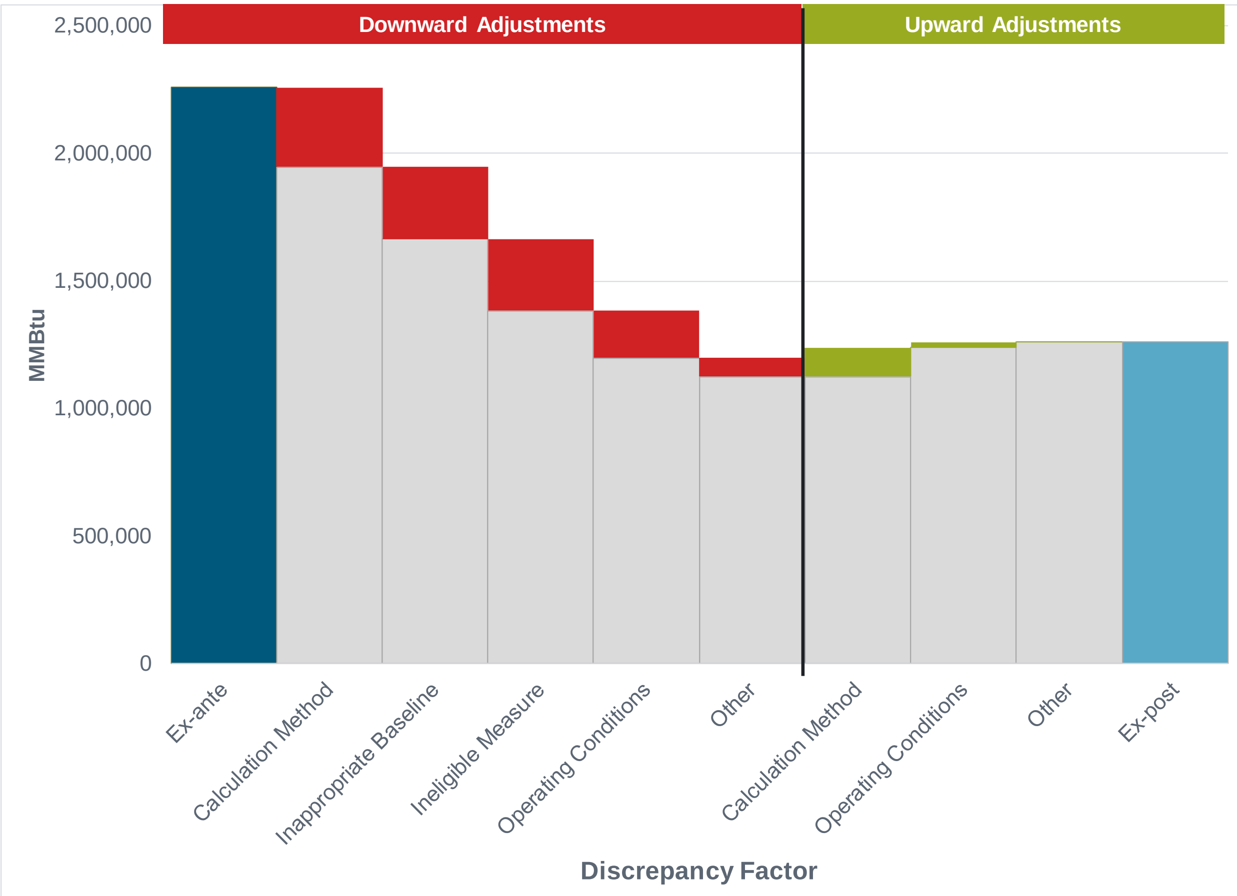
[18] In this case, more than 30 percent of sampled points had LC GRRs at the extreme tails (zero or less and greater than one), which results in a large standard deviation.

[19] 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.

[20] 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.

(ineligible measures), and eight percent (operating conditions).[21] Figure 2-3 depicts, by discrepancy factor, all downward and upward adjustments that were made to the ex-ante first year MMBtu savings estimates for the sample of 2015 M&V points.

FIGURE 2-3: SUMMARY OF DISCREPANCY FACTORS RESULTING IN DOWNWARD AND UPWARD ADJUSTMENTS TO FIRST YEAR EX-ANTE MMBTU IMPACTS - ALL PAS

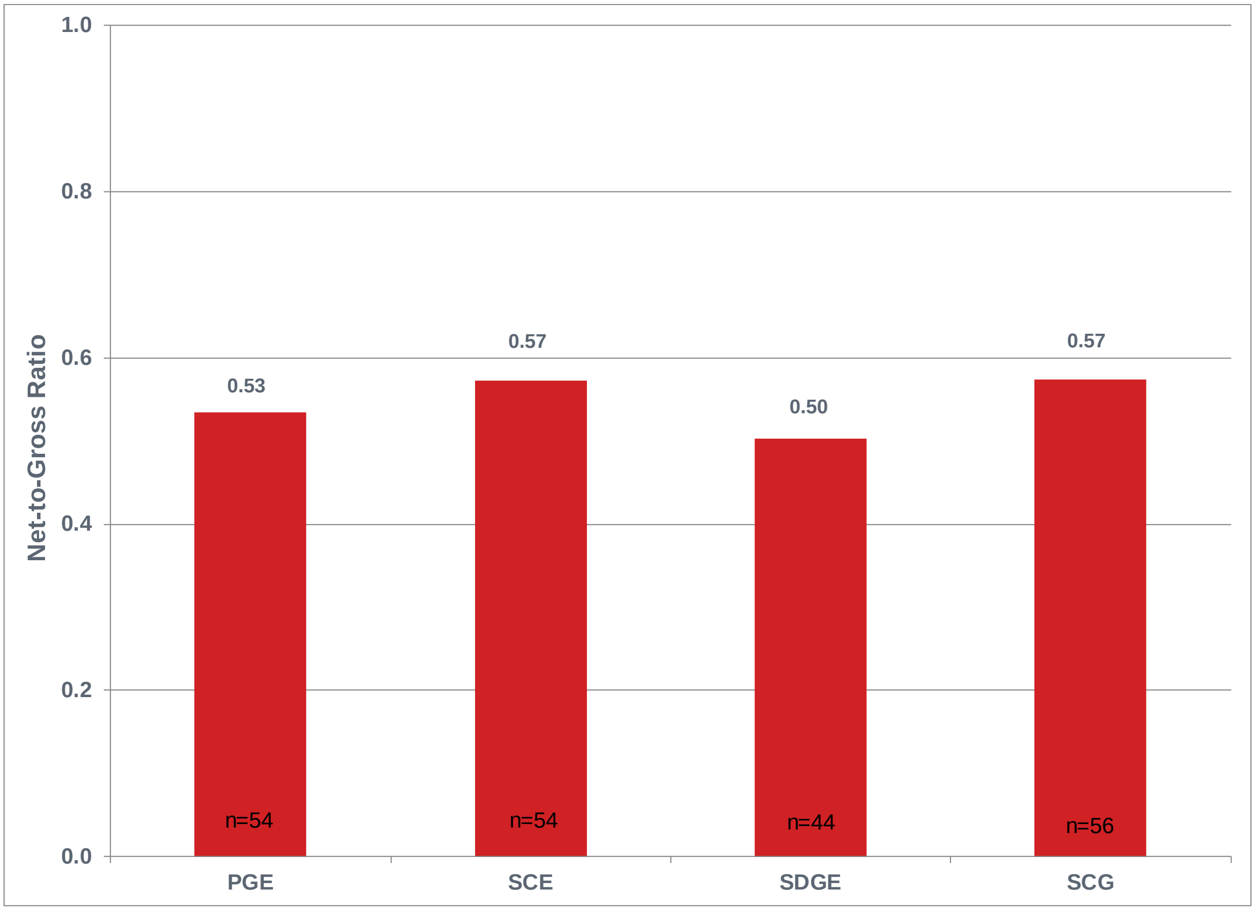


[21] Figures do not sum due to rounding.

2.3 HIGH LEVEL CUSTOM NET-TO-GROSS RESULTS

NTGR results at the PA level are presented in Figure 2-4 and Table 2-3.

FIGURE 2-4: WEIGHTED NET-TO-GROSS RATIOS BY PA[22]



[22] Note that these values reflect the removal of 13 projects from NTG calculations. As described in Chapter 5, this was due to either an ineligible measure (8 projects removed) or inappropriate baselines (5 project removed).

TABLE 2-3: WEIGHTED NET-TO-GROSS RATIOS BY PA

	Mean Net-to-Gross Ratios			
Results	**PGE**	**SCE**	**SDG&E**	**SCG**
Weighted NTGR	**0.53**	**0.57**	**0.50**	**0.57**
90 Percent Confidence Interval	0.49 to 0.58	0.53 to 0.61	0.44 to 0.57	0.54 to 0.61
Relative Precision	0.08	0.07	0.13	0.07
n NTGR Completes	54	54	44	56
N Sampling Units	1,351	765	98	196
Error ratio (ER)*	0.35	0.34	0.68	0.37
Sample Percent Representation of First Year Ex-Ante MMBtu Savings	29%	42%	64%	66%

* A measure of the statistical variation in the gross realization rates. Note that this error metric only captures sampling error, not measurement error.

Based on the NTGR results presented above and results from past evaluation cycles, the following observations are noteworthy:

- At the level of PA sampling domain, the final NTGRs range from 0.50 to 0.57, which are generally within the range of scores seen in the 2010-12 through 2014 evaluations (see weighted MMBtu comparison in Chapter 5).
- **PG&E:** The 2015 mean weighted NTGR result for PG&E (0.53) increased by 6 percent compared to 2014 but no statistically significant differences are observed across the past four evaluation cycles.
- **SCE:** The 2015 mean NTGR result for SCE (0.57) is the same as the 2013 result (0.57) but is higher than 2010-12 result (0.49) and 2014 result (0.46).
- **SDG&E:** The 2015 mean NTGR result for SDG&E (0.50) is lower than the 2013 result (0.59) but is similar to 2010-12 and 2014 NTGRs.
- **SCG:** For SCG the 2015 weighted mean NTGR across all projects is 0.57. SCG NTG scores have declined over the last three evaluation cycles but are still higher than the 2010-12 score of 0.49.

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.

This analysis was performed for the Lowest and Highest NTG ratio quartiles,[23] i.e., the group with the lowest NTGRs, corresponding to a threshold value of 0.36 and lower in 2015, and the group with the highest NTGRs, corresponding to a threshold value of 0.70 and higher in 2015. The goal of this analysis was to highlight the factors and characteristics of the groups of projects with both the strongest and weakest levels of program influence. This information not only leads to improved understanding of the results and underlying factors, but also highlights the characteristics of the strongest and weakest groups of projects with respect to program influence. In turn, these characteristics support the development of a set of actionable recommendations regarding strategies to improve program influence going forward.

The most important factors that differentiate those in the Highest NTG quartile are the program rebate (considered important by close to 100% of respondents), and program technical assistance or studies (considered important by nearly two-thirds of respondents). Key factors that distinguish those in the Lowest NTG quartile are the presence of industry standard practice, corporate policies favoring energy efficiency, and following normal maintenance and equipment replacement policies. These specific non-program motivations often drive project decision making for projects with Low program influence.

Another differentiating factor between the two groups was related to the timing of the program's involvement with the customer. The percentage of those that had made their decision before having any discussions with the program was close to zero for High NTG respondents, and nearly 90% (2015) for Low NTG respondents.

2.4 NET EVALUATION REALIZATION RATE RESULTS

Net evaluation realization rates are presented for each PA in Table 2-4 through Table 2-7. 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 gross ex-ante estimates by RR=0.9 as ordered by the CPUC in Decision 11-07-030. A total of 28 EAR projects were part of the IALC 2015 population: none were installed in PG&E territory, three in SCE territory, one in SDG&E territory, and 24 in SCG territory. This explains why the claimed GRR

[23] Each quartile consists of 25 percent of project results, and the groupings are assigned based on the NTGR values for the associated projects. Each quartile has identical numbers of projects within a given year.

from line b. in the tables that follow is higher than 0.90 in some cases. Two of these 28 projects were randomly sampled and were analyzed by the IALC Custom evaluation (refer to Chapter 4 for results).

TABLE 2-4: 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	2,440,885,949	345,238	135,941,362
b. Ex-Ante GRR	0.90	0.90	0.90
c. Ex-Ante Adjusted Gross Savings (c = a x b)	2,196,797,355	310,714	122,347,226
d. Ex-Ante NTGR	0.71	0.71	0.70
e. Claimed Net Savings (e = c x d)	1,563,963,677	221,762	85,174,196
f. Ex-Ante GRR x NTGR (f = b x d)	0.64	0.64	0.63
Evaluation			
g. Evaluation LC GRR	0.47	0.50	0.47
h. Evaluated Gross Results (h = a x g)	1,147,216,396	172,619	63,892,440
i. Evaluation NTG Ratio	0.53	0.53	0.53
j. Evaluated Net Results (j = h x i)	608,024,690	91,488	33,862,993
k. Evaluation GRR x NTGR (k = g x i)	0.25	0.27	0.25
l. Evaluated Net Realization Rate (l = j / e)	0.39	0.41	0.40

TABLE 2-5: SCE LIFECYCLE NET REALIZATION RATE ESTIMATES AND COMPARISONS

Impact Element	LC Electric Savings		LC Gas Savings[24]
	kWh	Avg. Peak kW	Therms
Tracking			
a. Ex-Ante LC Gross Savings	1,834,469,923	287,295	549,317
b. Ex-Ante GRR	0.90	0.90	0.90
c. Ex-Ante Adjusted Gross Savings (c = a x b)	1,652,307,245	258,693	494,385
d. Ex-Ante NTGR	0.68	0.68	0.45
e. Claimed Net Savings (e = c x d)	1,119,485,992	175,034	222,343
f. Ex-Ante GRR x NTGR (f = b x d)	0.61	0.61	0.40
Evaluation			
g. Evaluation LC GRR	0.41	0.40	0.41
h. Evaluated Gross Results (h = a x g)	752,132,668	114,918	225,220
i. Evaluation NTG Ratio	0.57	0.57	0.57
j. Evaluated Net Results (j = h x i)	428,715,621	65,503	128,375
k. Evaluation GRR x NTGR (k = g x i)	0.23	0.23	0.23
l. Evaluated Net Realization Rate (l = j / e)	0.38	0.37	0.58

TABLE 2-6: 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	267,068,305	39,310	18,027,780
b. Ex-Ante GRR	0.90	0.90	0.90
c. Ex-Ante Adjusted Gross Savings (c = a x b)	240,373,935	35,380	16,225,002
d. Ex-Ante NTGR	0.60	0.60	0.53
e. Claimed Net Savings (e = c x d)	145,169,630	21,201	8,605,371
f. Ex-Ante GRR x NTGR (f = b x d)	0.54	0.54	0.48
Evaluation			
g. Evaluation LC GRR	0.47	0.73	0.47
h. Evaluated Gross Results (h = a x g)	125,522,103	28,696	8,473,057
i. Evaluation NTG Ratio	0.50	0.50	0.50
j. Evaluated Net Results (j = h x i)	62,761,052	14,348	4,236,528
k. Evaluation GRR x NTGR (k = g x i)	0.24	0.37	0.24
l. Evaluated Net Realization Rate (l = j / e)	0.43	0.68	0.49

[24] SCE savings claims are largely associated with the Savings by Design Program, but also several energy efficiency partnership programs. Most of the tracking system records with gas savings claims are associated with whole-building new construction projects, but a few records involve MBCx measures, among others.

TABLE 2-7: SCG LIFECYCLE NET REALIZATION RATE ESTIMATES AND COMPARISONS

Impact Element	LC Gas Savings
	Therms/year
Tracking	
a. Ex-Ante LC Gross Savings	92,561,519
b. Ex-Ante GRR	0.91
c. Ex-Ante Adjusted Gross Savings (c = a x b)	83,793,680
d. Ex-Ante NTGR	0.50
e. Claimed Net Savings (e = c x d)	41,519,419
f. Ex-Ante GRR x NTGR (f = b x d)	0.45
Evaluation	
g. Evaluation LC GRR	0.51
h. Evaluated Gross Results (h = a x g)	47,206,375
i. Evaluation NTG Ratio	0.57
j. Evaluated Net Results (j= h x i)	26,907,634
k. Evaluation GRR x NTGR (k = g x i)	0.29
l. Evaluated Net Realization Rate (l = j / e)	0.65

2.5 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 Project Practices Assessment (PPA) activities. Extensive reporting of findings and recommendations is presented in Chapter 7 of this report. Readers are encouraged to examine Chapter 7 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,[25] including a review of evaluation EUL conclusions/rationale in an effort to improve EUL estimates and LC GRR results;

[25] It is notable that the evaluation estimate of EUL differed from the PAs tracking database estimate 52 percent of the time. For those instances the evaluation-derived average EUL was smaller than the ex-ante average EUL by roughly 3 years. As noted in Chapter 4, LC GRR results are lower than FY GRR results and this EUL difference is *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.

To reduce continued moderate free ridership, PAs should consider changes to program implementation procedures and features designed to increase program influence. These include: adopting procedures to identify and affect projects with low program influence; adjusting the set of technologies eligible for incentives; and implementing procedures to limit known free riders by upselling to higher efficiency levels, multi-measure solutions and continuous energy improvement.

Finally, key recommendations discussed in Chapter 7 of this report are listed in Table 2-8. 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.

TABLE 2-8: 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 PAs should 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)
Baseline Conditions
Increase efforts to ensure conformance with CPUC baseline policies and make a greater effort to examine existing equipment RUL
Clearly identify project event in terms of natural replacement, replace on burnout, early replacement, new construction, add-on equipment, and system optimization, and set the appropriate baseline accordingly
Appropriate interpretation and application of code requirements is needed, including the need to consider and possibly examine a broad array of codes and requirements that may be relevant for a given project

Key Recommendations by Topic Area

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, select a representative baseline period, and thoroughly document the pre-existing conditions for the purposes of establishing baseline
To improve project eligibility screening the PAs should 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, production records, and model inputs
Calculated savings should be based on robust data sets representing longer-term and stable operation of equipment and systems
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
The PAs should carefully review each of the 30 Final Site Reports (FSRs) listed in Section 4.4.2, Table 4-6, to identify the specific reasons that led zero or negative savings, and use those lessons learned to improve related project practices
A statewide document, similar to the PPA form, should be developed for use by all PAs for custom claims
The PA's project eligibility treatment suggests that the PA's communication and coordination efforts for disseminating, implementing and overseeing CPUC guidance should be improved

Key Recommendations by Topic Area
The PAs should prioritize M&V reviews for all large projects to ensure that CPUC M&V standards are being met for ex-ante savings estimation – a small number of projects normally account for half of all savings claims
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

3 INTRODUCTION AND BACKGROUND

This report presents draft results from the impact evaluation of the 2013-2015 California Program Administrator's (PAs)[26] energy efficiency programs, focusing on nonresidential custom measures installed during program year 2015 (PY2015). 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.[27] The evaluation is guided by the IALC Custom Impact Evaluation Plan dated November 2014.[28] The *IALC Custom Impact Research Plan PY2015 Addendum* dated April 2016, and the *IALC Research Plan PY2015 Sample Design Update* dated July 2016, provide additional detail on the evaluation effort; these evaluation plans are available on the ED public documents website.[29] Readers may also want to familiarize themselves with a number of other relevant CPUC sources that are referenced throughout this report.[30] This includes a nonresidential Net-to-Gross (NTG) methods document,[31] which can also be found on the ED public documents website.[32] 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.

---

[26] 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.

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

[28] http://www.energydataweb.com/cpucFiles/pdaDocs/1198/IALC_Research_Plan_Final_11-12-2014.pdf

[29] http://www.energydataweb.com/cpucFiles/pdaDocs/1541/2015%20Custom%20Research%20Plan%20Addendum.docx
http://www.energydataweb.com/cpucFiles/pdaDocs/1588/2015%20Custom%20%20Research%20Plan%20Addendum_JulySample_Final.pdf

[30] 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.

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

[32] The NTG methods document was distributed and discussed with PA project coordination group (PCG) and evaluation staff during previous evaluation efforts, starting in 2011.

3.1 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 a sample of 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 PY2013-2015 utility programs[33] 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.[34]

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 2015 claim period; the IALC 2013 and IALC 2014 reports are available on the ED public documents website.[35]

The CPUC organized all of its consultant evaluation and research work for PY2013-2015 into roadmaps.[36] 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

---

[33] In 2015 PG&E had 63 programs that include custom projects, SCE had 48 such programs, SDG&E had 6, and SCG had 6.

[34] 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.

[35] http://www.energydataweb.com/cpucFiles/pdaDocs/1341/IALC%202013%20Report%20Final%20071715.pdf
http://www.energydataweb.com/cpucFiles/pdaDocs/1518/IALC%202014%20Final%20Report%20April%202016.pdf

[36] See the 2013-2017 Energy Division & Program Administrator Energy Efficiency Evaluation, Measurement and Verification Plan Version 7, available at http://www.energydataweb.com/cpuc/home.aspx under the link for Energy Efficiency EM&V Plans.

group. Measure groups were then mapped and assigned to different roadmaps, each of which has its own project team, scope, and reporting. Mapping of assignments to road maps was also informed by residential versus nonresidential participation, deemed versus non-deemed (i.e., custom), upstream versus downstream provision of incentives, and other considerations. The IALC roadmap was assigned all of the nonresidential custom projects, excluding lighting and codes and standards claims.[37]

Energy savings claims from the measures assigned to the IALC roadmap represent a significant contribution to the overall savings portfolios for the PAs' 2015 energy efficiency programs, accounting for 20 percent of statewide electric savings claims and 43 percent of statewide positive gas savings claims. In 2015 the PA tracking data for measures assigned to the IALC roadmap included thousands of entries statewide with annual adjusted gross ex-ante electric savings by the PAs totaling 355 GWh and 50 MW.[38] Statewide PA positive adjusted gross ex-ante gas savings for measures assigned to the IALC roadmap total 18 million Therms.

California PA-led custom programs are currently evaluated on an annual basis. The majority of the evaluation field work and data collection activities took place in the second half of 2016, and the analysis and reporting was completed in the first half of 2017. 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.[39]

The most recent PA data extract, which reflects cumulative PA program activity through the fourth quarter of 2015,[40] was used to summarize the 2015 claimed energy savings associated with the PA portfolios, as well as the savings assigned to the IALC roadmap's custom impact evaluation.[41] These savings are reported in Table 3-1.

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

38 Adjusted gross ex-ante savings are calculated as gross ex-ante savings times the 0.9 RR custom measure adjustment ordered by the CPUC in D.11.07.030.

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

40 Savings in the Q4, 2015 database were frozen as of January 6, 2017.

41 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 3-1: 2015 FIRST YEAR ADJUSTED GROSS EX-ANTE ENERGY IMPACTS BY PA FOR PROJECTS IN THE IALC ROADMAP, AND IN THE PORTFOLIO

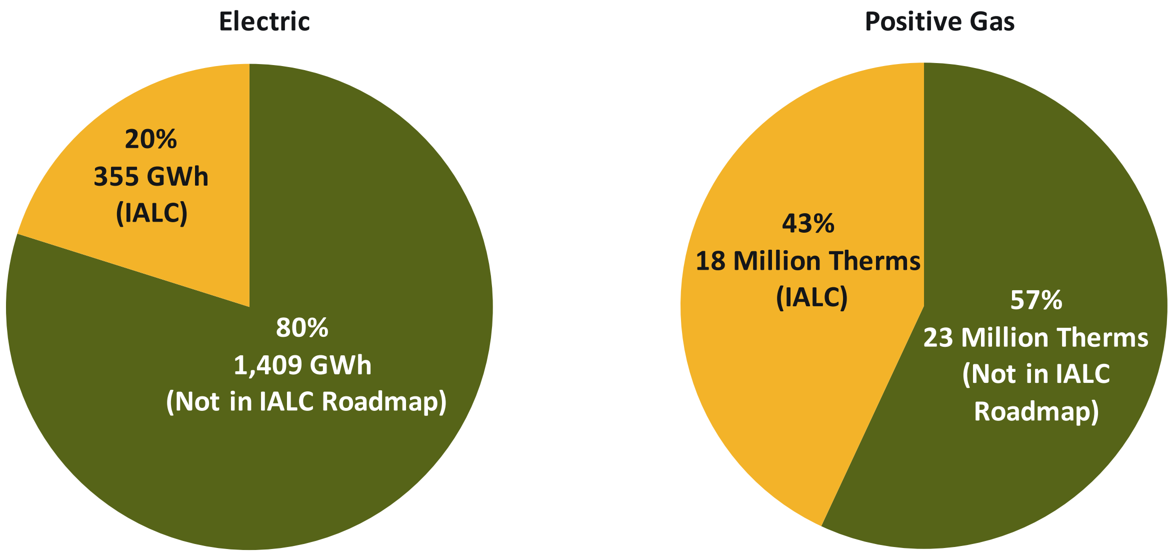
FY Adjusted Gross Ex-Ante Impacts by PA			
PA	**Electric Energy (GWh)**	**Electric Demand (MW)**	**Gas Energy (Million Therms)[42]**
2015 PA Adjusted Gross Ex-Ante Savings, IALC Roadmap			
PG&E	182	24	10
SCE	154	24	0
SCG	0	0	7
SDG&E	20	3	1
Total*	355	50	18
Total 2015 PA Adjusted Gross Ex-Ante Savings			
PG&E	733	144	22
SCE	857	155	2
SCG	14	7	15
SDG&E	161	32	3
Total*	1,764	338	41
IALC Percentage of Total PA Adjusted Gross Ex-Ante Savings			
PG&E	25%	16%	45%
SCE	18%	15%	5%
SCG	0%	0%	45%
SDG&E	12%	8%	48%
Total*	20%	15%	43%

* Totals may not sum due to rounding.

[42] 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. These records were not included in the sample population and sample design for gas projects, as their inclusion would otherwise allow a sample that would not accurately represent the actual number and aggregate savings of the natural gas projects implemented. The IALC evaluation addresses lighting only in conjunction with other measures that are all part of a new construction whole building project. Negative Therm records were therefore not included in the Table 3-1 energy saving claims summary.

Figure 3-1 contrasts the IALC roadmap first year adjusted gross ex-ante savings with total portfolio first year adjusted gross ex-ante savings for 2015.

FIGURE 3-1: 2015 IALC ROADMAP FIRST YEAR ADJUSTED GROSS EX-ANTE SAVINGS RELATIVE TO PORTFOLIO SAVINGS



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, starting in PY2014 and continuing in 2015, 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, one additional group of projects emerged. Through its Agriculture Energy Advisor Program (SCE-13-SW-004A), SCE customers can benefit from full service pump efficiency improvement services (a.k.a. agricultural pump testing). This measure was evaluated in the 2006-2008 program cycle and is not currently in scope for the 2015 evaluation. The 3,488 records corresponding to SCE pump testing that were assigned to the IALC roadmap in 2015 were not evaluated. Table 3-2 shows how the records assigned to the IALC roadmap in 2015 were further separated into Pump Testing, NRNC, and Custom groups. Only positive gas claims are shown.

TABLE 3-2: 2015 FIRST YEAR AJDUSTED GROSS EX-ANTE ENERGY IMPACTS FOR THE IALC ROADMAP, BY GROUP AND PA

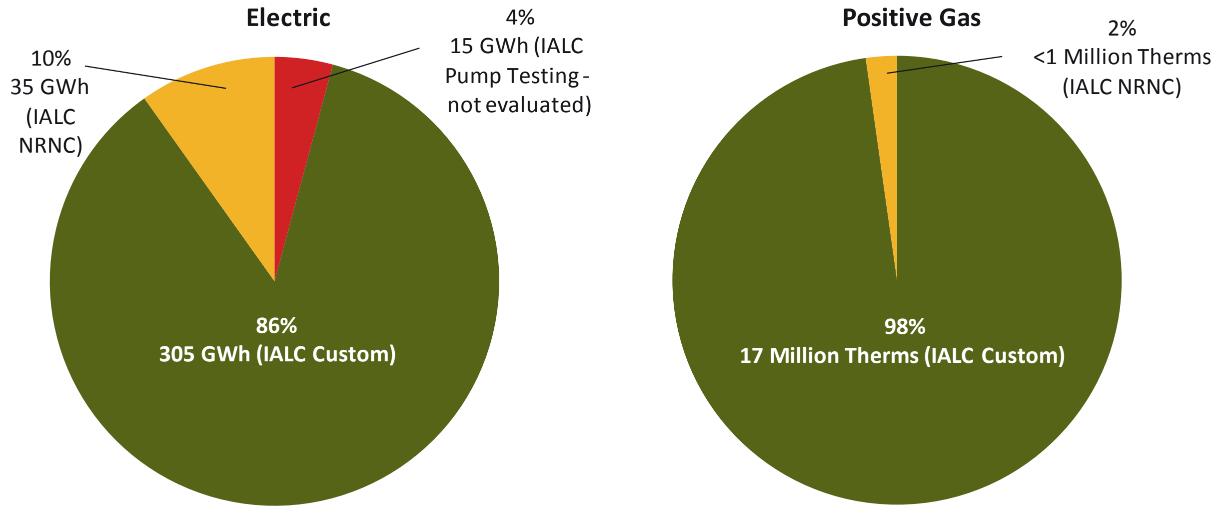
FY Adjusted Gross Ex-Ante Impacts by PA

PA	Electric Energy (GWh)	Electric Demand (MW)	Positive Gas Energy (Million Therms)
IALC 2015 PA Adjusted Gross Ex-Ante Savings			
PG&E	182	24	10
SCE	154	24	0
SCG	0	0	7
SDG&E	20	3	1
Total*	355	50	18
IALC 2015 Pump Testing PA Adjusted Gross Ex-Ante Savings - Not Evaluated			
PG&E	0	0	0
SCE	15	3	0
SCG	0	0	0
SDG&E	0	0	0
Total*	15	3	0
IALC 2015 Whole Building NRNC PA Adjusted Gross Ex-Ante Savings - Evaluated			
PG&E	22	4	0
SCE	9	3	0
SCG	0	0	0
SDG&E	4	1	0
Total*	35	8	0
IALC 2015 Custom Savings Adjusted Gross Ex-Ante - Evaluated			
PG&E	160	19	9
SCE	129	18	0
SCG	0	0	7
SDG&E	16	1	1
Total*	305	39	17

* Totals may not sum due to rounding.

Figure 3-2 presents the Custom and NRNC portions of the IALC roadmap adjusted gross ex-ante savings that were included in the 2015 evaluation. The figure also shows the Pump Testing portions that are not in scope in 2015. Only positive gas savings are shown.

FIGURE 3-2: 2015 FIRST YEAR ADJUSTED GROSS EX-ANTE SAVINGS ASSIGNED TO THE IALC ROADMAP



Please note that Table 3-1, Table 3-2, Figure 3-1 and Figure 3-2 reflect PA *claimed* gross savings (adjusted gross ex-ante savings), i.e., PA gross savings claims adjusted by the 90 percent default PA GRR for all records not affected by the EAR process.[43] **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 GRR is backed out (i.e., PA savings claim / 0.9 = gross unadjusted savings estimate). In addition, only projects with a positive MMBtu ex-ante savings estimate were evaluated. MMBtu describes combined electric and gas savings. Refer to Table 3-3 for total IALC PY2015 unadjusted gross savings estimates for all projects evaluated.

[43] 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 PAs' 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 GRR (in other words: PA GRR=1.0 for these projects).

Evidence suggests that some projects that have been "touched" by the EAR process do not always have PA GRR = 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 GRR 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 GRR equal to 1.0, the evaluation uses this tracking data-based source to identify EAR points (PA GRR=1.0).

TABLE 3-3: 2015 UNADJUSTED FIRST YEAR GROSS SAVINGS ESTIMATES EVALUATED BY IALC CUSTOM, BY PA

PA	Electric Energy (GWh)	Electric Demand (MW)	Gas Energy (Million Therms)
IALC 2015 Evaluated Unadjusted FY Gross Savings			
PG&E	202	26	11
SCE	154	23	0
SCG	0	0	7
SDG&E	22	3	1
Total	378	53	19

3.2 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.[44]
3. Providing timely feedback to PAs to facilitate program design improvements.
4. Determining whether the impact estimation methods, inputs, and procedures used by the PAs and implementers are consistent with the CPUC's policies, decisions and best practices.[45]

---

[44] The IALC Custom NTG surveys also include a battery of questions to address participant spillover. However, estimation of spillover is not part of the IALC scope of work. For 2013-2014, these data were 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

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

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 148 gross impact points and 208 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 4. 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 5.

In addition, Project Practices Assessments (PPAs) were incorporated into the 2013-2015 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. The PPA process was based on all sampled gross impact points, and among other objectives seeks evidence of EAR influence on ex-ante site-level M&V protocols, procedures and results.[46] The results of the PPA analysis are discussed in Chapter 6.

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 2013-2014 custom impact evaluations), 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 2015 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.

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.[47] 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.

[46] 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.

[47] 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.

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.

3.3 STRUCTURE OF THE REPORT

Table 3-4 shows the overall organizational structure of this report. Although overarching findings and recommendations are in Chapter 7, it is noteworthy that findings are also highlighted in Chapters 4, 5 and 6. 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 3-4: OVERALL ORGANIZATIONAL STRUCTURE OF REPORT

Section #	Title	Content
1	**Executive Summary**	High level summary of the study and results
2	**Technical Summary**	Summary of results and key findings
3	**Introduction and Background**	Evaluation objectives, research issues, and savings claims
4	**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
5	**NTG Results**	Net of free ridership ratios and results, and key factors influencing NTGRs
6	**PPA Results**	Assessments based on a comparison between ex-ante and ex-post M&V-based conclusions
7	**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

4 GROSS IMPACT RESULTS

This chapter presents quantitative and qualitative gross impact results for the 2015 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)[48] domain and size stratification segment. Results are also presented for energy metrics – source energy (MMBtu)[49] and electric demand (kW).

Stratified sampling was implemented for custom measures installed in 2015 by each PA separately: PG&E, SCE, SDG&E and SCG (for more detail please refer to the Custom Evaluation Research Plan, PY2015 Addendum and PY2015 Sample Design Update referenced in Chapter 3). 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.[50]

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[51]
- Approved List of ISP Studies[52]

---

[48] 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.

[49] 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.

[50] 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.

[51] Industrial, Agricultural and Large Commercial Evaluation Guidance from IALC_2 WO available at www.energydataweb.com/cpuc/. Select "advanced search," and from the drop down menus, select Work Order (ED_I_IAL_2-Itron). Then click the Search button. Direct link: http://www.energydataweb.com/cpucFiles/pdaDocs/1256/Evaluation%20Guidance%20Questions%20for%20Site%20Specific%20Analysis_2014_0918.pdf

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

4.1 PROJECT-SPECIFIC GROSS IMPACT SUMMARY

Gross impact evaluation results are supported by 148 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 148 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 2015 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 146 main points and 2 backup points, collectively representing 41% of ex-ante lifecycle gross MMBtu savings estimates across all PAs (Table 4-1). Backup points serve to fill-in for main points where completion of main points is not possible.

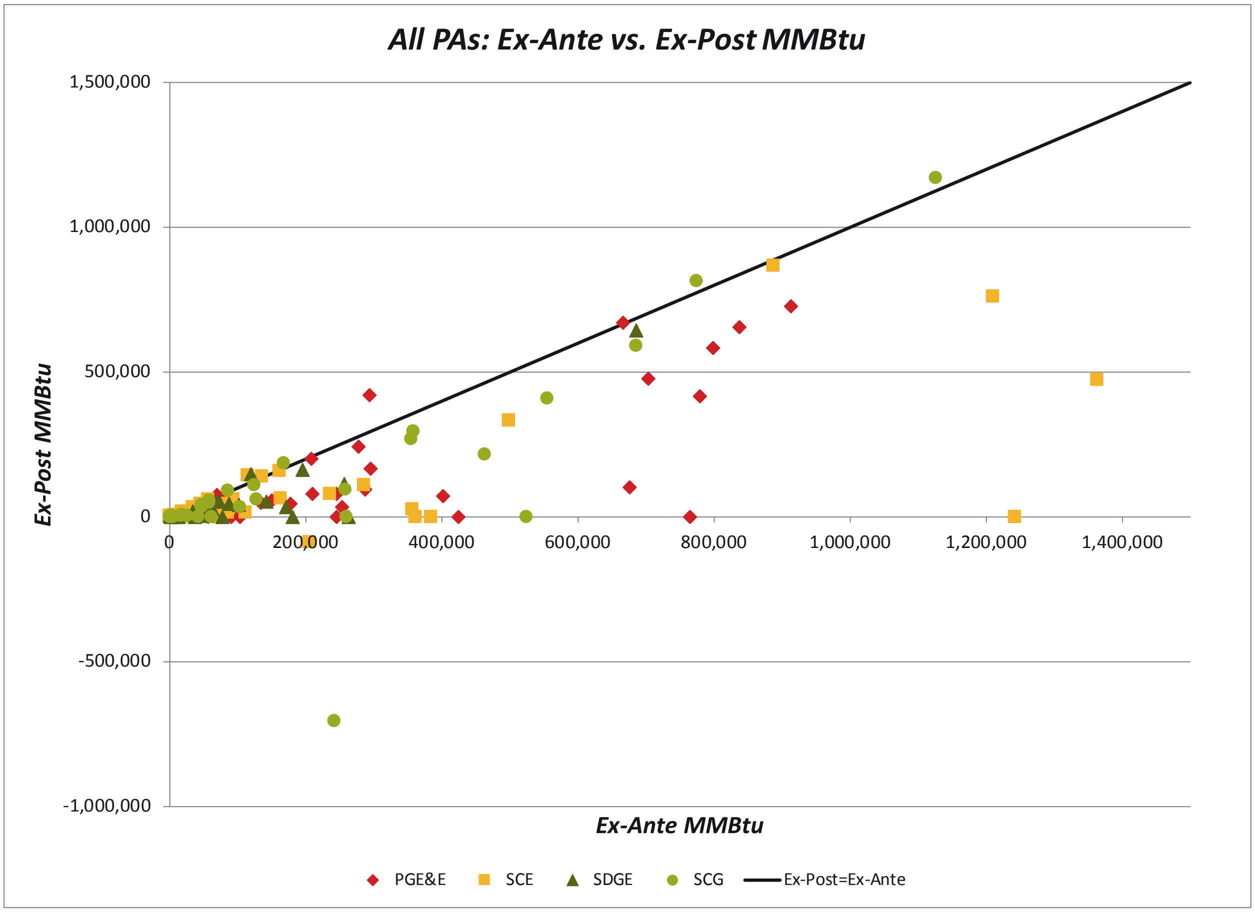
TABLE 4-1: CUSTOM EVALUATION GROSS M&V SAMPLE DISPOSITION BY PA

PA	Completed M&V Points (n)		Percent of Ex-Ante LC MMBtu Gross Savings Estimates
	Main	Backup	M&V Sample
PG&E	42	0	28%
SCE	42	1	47%
SDG&E	32	1	64%
SCG	30	0	72%
All PAs	**148**		**41%**

Figure 4-1 graphically displays MMBtu-based gross ex-post versus gross ex-ante *lifecycle* savings estimates for the statewide sample. The figure compares the gross ex-ante (tracking system) MMBtu savings[53] with the gross ex-post evaluated MMBtu savings for each of 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). 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.

[53] 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.

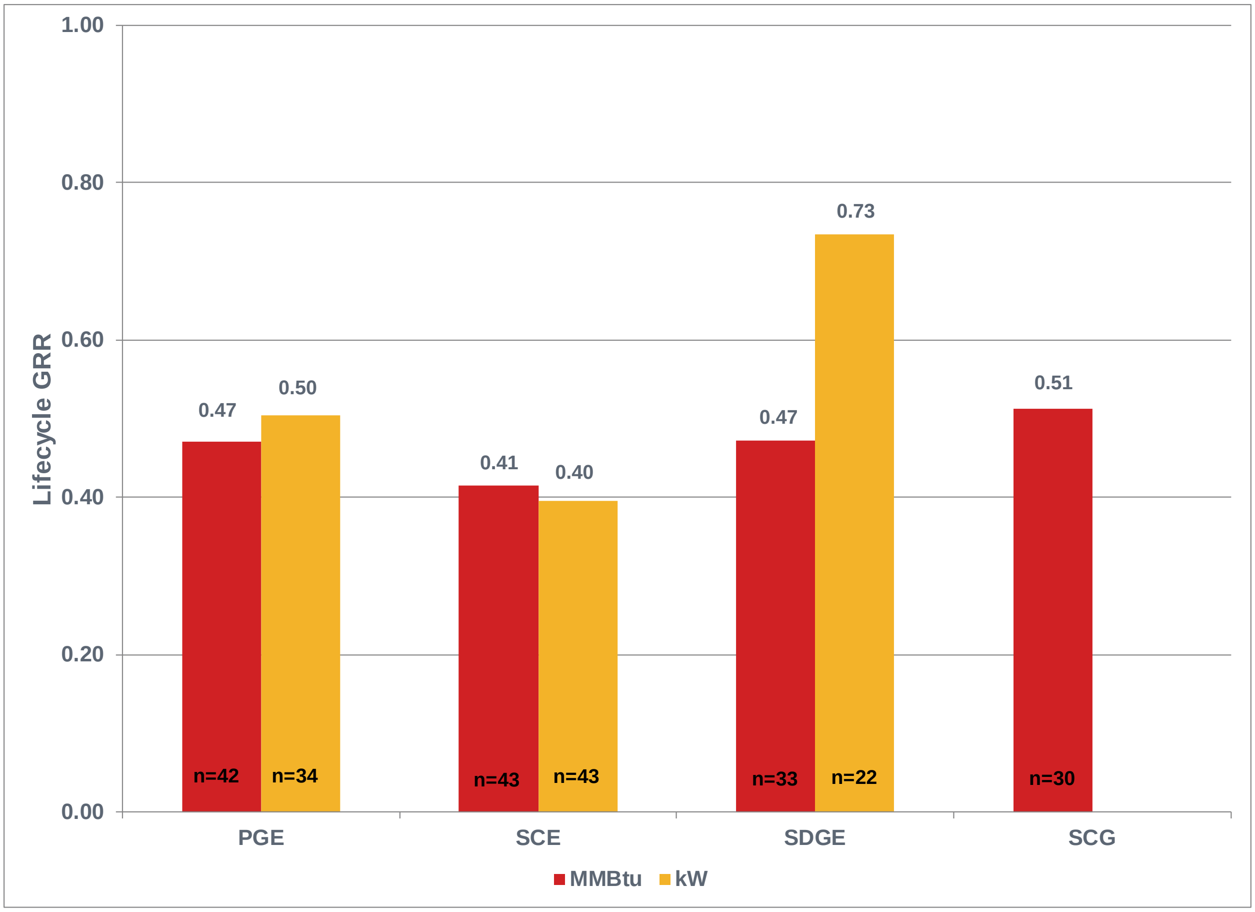
FIGURE 4-1: LIFECYCLE EX-ANTE AND EX-POST COMBINED ELECTRIC AND GAS GROSS SAVINGS (MMBTU) FOR SAMPLED PROJECTS



4.2 PA GROSS REALIZATION RATE RESULTS

Weighted gross realization rates by PA and energy metric (MMBtu or kW) are presented graphically in Figure 4-2.

FIGURE 4-2: 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 4-2 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.[54] In addition, first year (FY) GRRs are presented for comparison purposes.

[54] The error ratio is a measure of the statistical variation in the gross realization rates. Note that this error metric only captures sampling error, not measurement error.

TABLE 4-2: WEIGHTED PROJECT LIFECYCLE AND FIRST YEAR GROSS REALIZATION RATES BY PA AND ENERGY METRIC (MMBTU AND KW)

Energy Metric	Population Count	Sample Count	% of LC 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										
MMBtu*	1,351	42	28%	0.47	0.96	0.36 to 0.58	0	7	2	0.54
kW	1,066	34	20%	0.50	1.17	0.34 to 0.67	2	6	2	0.64
SCE										
MMBtu*	765	43	47%	0.41	0.73	0.34 to 0.49	0	5	3	0.55
kW	702	43	38%	0.40	1.32	0.27 to 0.52	2	7	5	0.50
SDG&E										
MMBtu*	98	33	65%	0.47	0.66	0.40 to 0.55	0	4	2	0.51
kW	67	22	49%	0.73	1.22	0.48 to 0.99	4	4	1	0.77
SCG										
MMBtu	196	30	72%	0.51	0.92	0.38 to 0.64	0	5	2	0.53

* 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 for all PAs, as not every project included a kW saving claim.

** A measure of the statistical variation in the gross realization rates. Note that this error metric only captures sampling error, not measurement error.

The error ratios for most domains are similar to the error ratios obtained in the *2010-2012 WO033 Custom Impact Evaluation Final Report,* the *2013 IALC Custom Impact Evaluation Final Report,* and the *2014 IALC Custom Impact Evaluation Final Report.*[55] In the 2015 sample, the relatively high standard deviation is largely 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 SCE's kW GRR consisted of 43 projects, two of which had LC GRRs of greater than 1.0 and 12 projects with LC GRRs of zero or less, resulting in an error ratio of 1.32.[56]

---

55 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
http://www.calmac.org/publications/IALC_2014_Final_Report_April_2016.pdf

56 In this case, more than 30 percent of sampled points had LC GRRs at the extreme tails (zero or less and greater than one), which results in a large standard deviation.

The mean lifecycle realization rates by PA and energy metric are less than 0.52 for all but one energy metric (the GRR for SDG&E kW is 0.73) and are similar, but generally lower, than those from the 2010-2012, 2013, and 2014 evaluations. As a comparison to 2015 results presented in the table above, Table 4-3, Figure 4-3 and Figure 4-4 include LC GRR evaluation results from the 2010-12, 2013, and 2014 cycles on a combined MMBtu basis.[57] LC GRR results appear in the upper half of Table 4-3. While few of the MMBtu-based LCC GRR results are statistically different at the 90% confidence level across the four evaluation cycles, 2015 MMBtu GRR results are lower than the past three cycles for every PA except SCG, which showed a small increase over 2014 (though still lower than results obtained in both the 2010-12 and 2013 evaluations).

Table 4-3 also includes a similar comparison of FY GRR results in the lower half of the table. Relative to the 2014 custom impact evaluation results PG&E, SCE, SDG&E and SCG FY MMBtu GRRs decreased by about nine percent, 14 percent, 30 percent, and eight percent, respectively. 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, on the other hand, are an indication of PA performance for all FY engineering elements plus EUL and early retirement (ER) treatment (including associated RUL and EUL considerations).

[57] 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, 2014, and 2015 results.

TABLE 4-3: 2010-2012, 2013, 2014, AND 2015 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	2015 Mean Gross Realization Rate	2015 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	0.47	0.36 to 0.58
kW	0.46	0.35 to 0.58	0.44	0.28 to 0.61	0.74	0.34 to 1.14	0.50	0.34 to 0.67
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	0.41	0.34 to 0.49
kW	0.57	0.47 to 0.67	0.52	0.43 to 0.62	0.46	0.34 to 0.58	0.40	0.27 to 0.52
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	0.47	0.4 to 0.55
kW	0.82	0.46 to 1.17	0.76	0.57 to 0.95	0.63	0.54 to 0.71	0.73	0.48 to 0.99
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	0.51	0.38 to 0.64
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	0.54	0.42 to 0.67
kW	0.53	0.41 to 0.66	0.54	0.37 to 0.70	0.69	0.28 to 1.10	0.64	0.46 to 0.83
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	0.55	0.47 to 0.63
kW	0.61	0.53 to 0.70	0.64	0.53 to 0.76	0.50	0.37 to 0.64	0.50	0.36 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	0.51	0.44 to 0.58
kW	0.84	0.48 to 1.19	1.02	0.88 to 1.17	0.67	0.52 to 0.81	0.77	0.52 to 1.02
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	0.53	0.43 to 0.63

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

FIGURE 4-3: COMPARISON OF 2010-12, 2013, 2014, AND 2015 WEIGHTED MMBTU LC GRR RESULTS

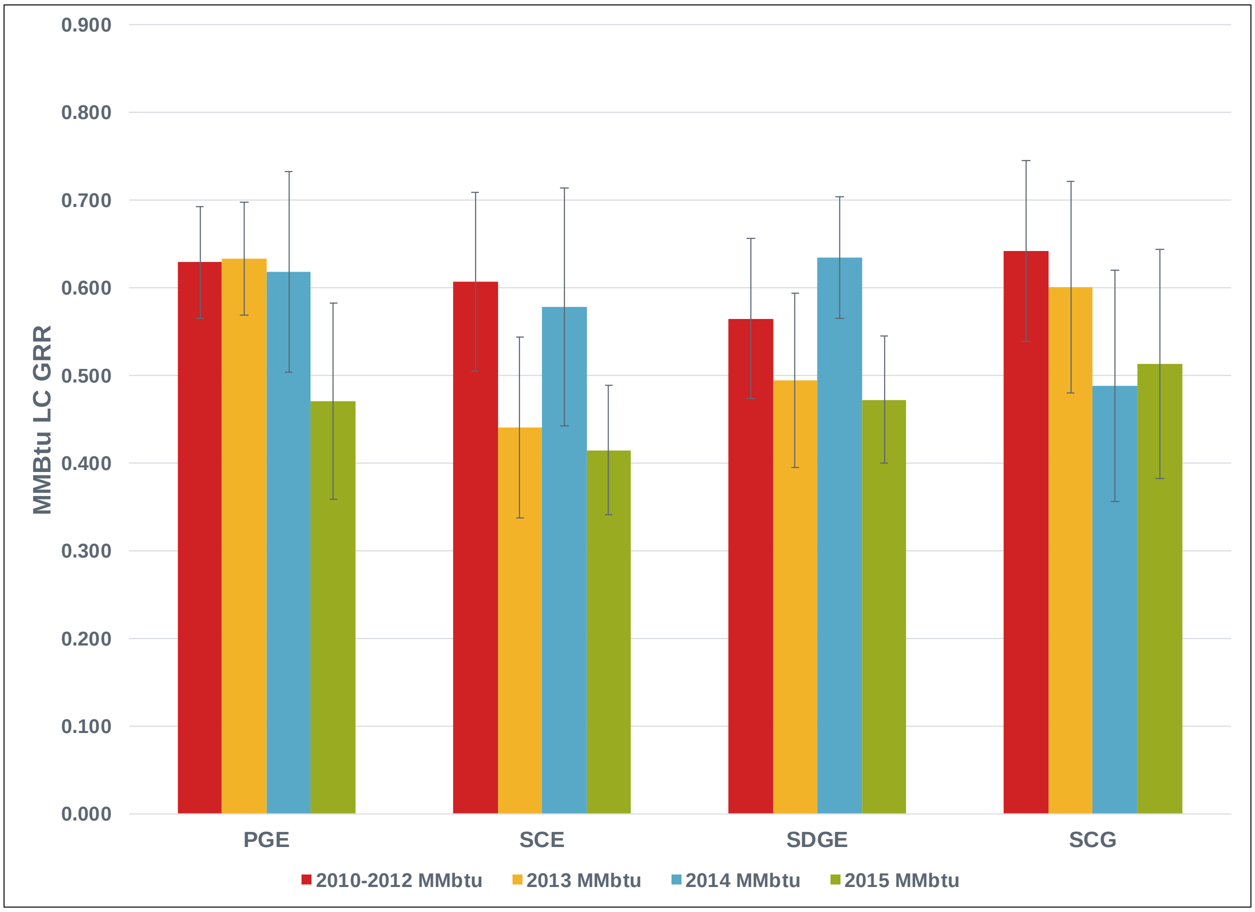
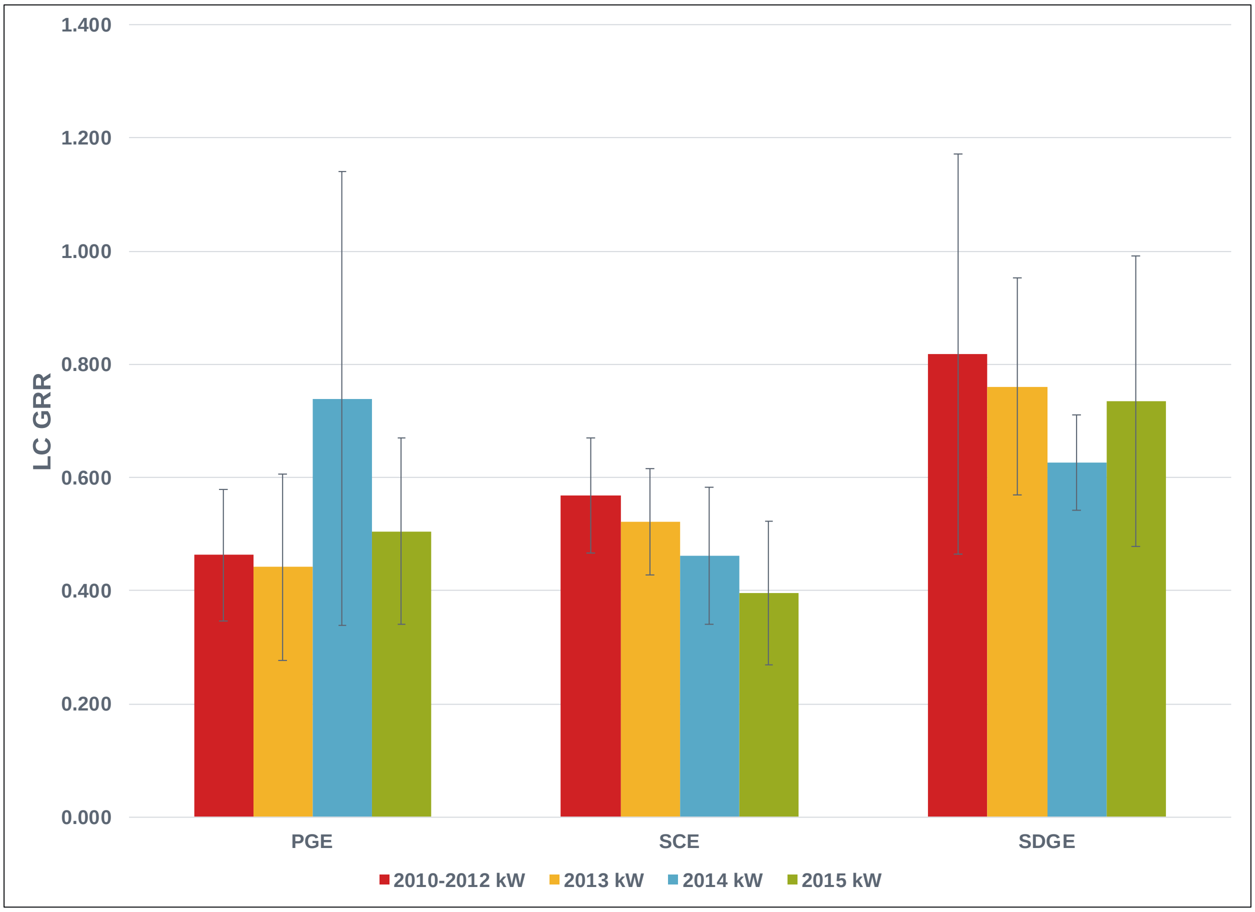


FIGURE 4-4: COMPARISON OF 2010-12, 2013, 2014, AND 2015 WEIGHTED KW LC GRR RESULTS



For all PAs, weighted lifecycle realization rates in 2015 are lower than the corresponding first year realization rate, which is primarily due to adjustments in measure effective useful life (EUL).[58] In particular, of the 186 measures (some of the 148 projects had multiple measures) evaluated in 2015, the ex-ante EUL was overstated in the tracking system extract for 71 measures (for example: a measure with 5-year life expectancy was assigned a 15-year EUL.) There were also 25 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 387 years compared to understatement of 82 years. Section 6.3.5 contains a more thorough assessment of ex-ante EUL assignments, as well as differences in comparison with ex-post EUL assignments.

[58] EUL is not included as a part of the "Discrepancy Analysis" analysis presented in Section 4.4 as that analysis is engineering focused and only examines first year impacts. The differences between FY and LC GRRs shown in Table 4-3 serve as a good proxy for quantifying the LC savings reductions resulting from differences in EUL.

All PAs had projects with negative and zero GRRs, and these served to lower the weighted average realization rate results. The discrepancy factors that brought about the lower realization rate results are explored in Section 4.4.

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 4-4 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.

TABLE 4-4: PROJECT LIFECYCLE REALIZATION RATES BY STRATA AND SAMPLE DOMAIN

PA Domain	Strata	Sample Count	% of LC MMBtu Savings Sampled	Projects with kW Ex-Ante	% of LC 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	0	-	0	-	-	-	-	-	-
	2	9	58%	7	91%	0.65	0.57	0	2	0
	3	10	43%	7	53%	0.26	0.48	0	1	0
	4	12	12%	10	15%	0.38	0.26	0	2	1
	5	11	2%	10	2%	0.62	0.54	0	2	1
SCE	1	6	100%	6	100%	0.45	0.43	0	1	1
	2	8	54%	8	45%	0.08	0.10	0	2	1
	3	10	30%	10	27%	0.83	0.70	0	1	0
	4	10	14%	10	9%	0.26	0.44	0	0	1
	5	9	1%	9	2%	0.36	0.38	0	1	0
SDG&E	1	1	100%	1	100%	0.66	0.60	0	0	0
	2	3	100%	2	100%			0	1	0
	3	8	62%	5	84%	1.04	0.46	0	0	1
	4	11	53%	8	39%	0.76	0.44	0	1	1
	5	10	9%	6	6%	0.50	0.27	0	2	0
SCG	1	0	-	-	-	-	-	-	-	-
	2	7	100%	-	-	-	0.78	0	1	0
	3	5	74%	-	-	-	-0.01	0	1	1
	4	9	45%	-	-	-	0.64	0	1	0
	5	9	12%	-	-	-	0.30	0	2	1

Observations on Table 4-4 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 4 and SCE's stratum 2). Similarly, a stratum containing large number of projects with GRRs at or above 1.0 will have a stratum-level GRR that is higher than the PA average (e.g. SCE's stratum 3 had four projects with MMBtu GRR >= 0.97 and SCG's stratum 2, which had two projects with MMBtu GRR > 1.0).
- High LC kW GRRs for SDG&E in strata 3 and 4 served to increase the overall weighted LC kW GRR for this PA.
- Low LC GRRs in SCE stratum 2 (both kW and MMBtu) and SCG LC MMBtu GRR stratum 3 brought the overall weighted averages for these PAs down. The negative LC MMBtu GRR for SCG stratum 3 results from one zero saver and one extremely negative saver (LC GRR = -2.9).
- SDG&E's stratum 3 LC kW GRR of 1.04 is driven primarily by one project with an extremely high GRR of 7.04.

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 4-6, 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.

4.3 EAR OVERLAP SENSITIVITY ANALYSIS

The establishment of ex-ante review (EAR) is discussed in CPUC Decision 09-09-047,[59] which requires the Energy Division (ED) to review and approve ex-ante impact estimation approaches and ex-ante savings for

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

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 PAs (in other words: PA GRR=1.0). To claim all other non-deemed projects, PAs adjust ex-ante estimates by a PA GRR=0.9.

A total of 28 EAR projects (PA GRR=1.0) were part of the IALC 2015 population: none were installed in PG&E territory, three in SCE territory, one in SDG&E territory, and 24 in SCG territory. The IALC stratified random sampling process selected two EAR projects for evaluation: one from SCE, one from SCG, and none from PG&E and SDG&E. Table 4-5 shows the first year and the lifecycle MMBtu realization rate results for these two 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[60] 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 purposes.[61]

---

[60] 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 GRR=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.

[61] Note that this constitutes a comprehensive sensitivity analysis only if all EAR-reviewed projects can be identified in the database by searching for PA GRR=1.0. For the purpose of this analysis, any projects identified as PA GRR=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 GRR = 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 GRR 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 GRR equal to 1.0, the evaluation uses this tracking data-based source to identify EAR points (PA GRR=1.0).

TABLE 4-5: MMBTU REALIZATION RATES FOR SAMPLED EAR PROJECTS

PA Domain	Project ID	PA GRR	FY GRR-MMBtu	LC GRR-MMBtu
PGE	All Sample (n=42)	0.90	0.54	0.47
	Percent change due to EAR points	-	-	-
SCE	F50013	1.00	0.93	0.35
	Remaining Points (n=42)	0.90	0.54	0.42
	All Sample (n=43)	0.90	0.55	0.41
	Percent change due to EAR points	0%	1%	-1%
SDG&E	All Sample (n=33)	0.90	0.51	0.47
	Percent change due to EAR points	-	-	-
SCG	G50003	1.00	0.75	0.75
	Remaining Points (n=29)	0.90	0.52	0.51
	All Sample (n=30)	0.91	0.53	0.51
	Percent change due to EAR points	1%	3%	1%

Observations for Table 4-5:

- The SCE sampled EAR point (F50013) has FY and LC MMBtu GRRs < 1.0. The seven percent (one minus 0.93) reduction observed for the FY GRR is due to operating conditions, while the LC GRR is driven down by a factor of 65 percent (one minus 0.35) due to a reduction in EUL (from eight to three years based on CPUC EAR staff recommendation).
- The SCG sampled EAR point (G50003) had reductions in FY and LC GRRs due to production changes observed during the ex-post M&V period.

Table 4-5 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 2015 IALC population (and sample) had minimal impact on the overall FY and LC MMBtu GRRs, ranging across energy metrics from negative one percent to three percent.

4.4 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 gross MMBtu impacts.[62]

When first year gross ex-post impacts for a sampled project were found to be different than the first year gross 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.[63] 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:[64]

- Calculation methods used for ex-post savings estimation were different than those used to estimate ex-ante savings. Some examples of ex-post methods include: running whole building SBD project simulations using the non-compliance mode to estimate savings and compliance mode to demonstrate project eligibility; 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; use of pre- and post-installation M&V, including measurement of calculation parameters, collection of relevant production records, and use of measured versus assumed inputs; 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; regressive baselines; new equipment that does not exceed code-, ISP- or regulation-required efficiency levels; and inaccurate baseline or pre-retrofit operating hours.

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

[63] A separate 'Other' category includes less common factors and generally accounts for a relatively small number of projects and percentage change in savings claims.

[64] While inappropriate baseline may ordinarily cause a downward adjustment (ex-post lower than ex-ante impacts), adjustments to the operating conditions, measure count, and/or calculation method sometimes caused an upward adjustment (where ex-post savings estimates are higher than ex-ante estimates).

- 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); and other program rule violations.
- Differences in operating conditions (for example, occupancy schedules, changes in hours of operation, VSD speeds, return to original operation, changes in production levels, etc.).[65]

4.4.1 Summary of Discrepancy Factor Impact

Given multiple tracking records associated with some projects, 188 records associated with the impact sample of 148 projects were examined (representing 2.3 Million MMBtu of gross ex-ante savings). For 14 records, the evaluation found no discrepancies (0.1 MMBtu of gross ex-ante savings were not adjusted). For the balance of 174 records and 2.2 MMBtu of gross ex-ante savings, gross ex-post estimates were different from gross 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 50 percent reduction in ex-ante savings estimates, while the upward discrepancies accounted for a six percent boost, resulting in a net downward adjustment of 44 percent. A summary of these adjustments is presented in this section.

4.4.2 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 22 projects for which the ex-post MMBtu impacts were zero, and eight for which the ex-post impacts for the project were negative. Table 4-6 identifies these projects and the factors that led to the zero or negative ex-post MMBtu impacts. Also shown is the extent of MMBtu reduction to ex-ante savings estimates for each project, by discrepancy factor.

For projects with zero gross 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 and operating conditions were the factors that occurred most frequently for projects with negative ex-post MMBtu.

[65] 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 may have changed before the time of the PA's or implementer's final inspection, but ex-ante savings were not always updated in such instances.

The evaluation conducted a sensitivity analysis to better understand the influence that these 30 projects had on the resulting weighted MMBtu-based LC GRR results. Removal of the nine PG&E projects resulted in a 20 percent increase in the LC GRR result, from 0.47 to 0.56. Removal of the eight SCE projects resulted in a 28 percent increase in the LC GRR result, from 0.41 to 0.53. Removal of the six SDG&E projects resulted in a 26 percent increase in the LC GRR result, from 0.47 to 0.60. Removal of the seven SCG projects resulted in a 46 percent increase in the LC GRR result, from 0.51 to 0.75.

A sensitivity analysis was also conducted to understand the influence that these 30 projects had on the resulting weighted FY GRR MMBtu results. Removal of the nine PG&E projects resulted in a 27 percent increase in the FY GRR result, from 0.54 to 0.69. Removal of the eight SCE projects resulted in a 28 percent increase in the FY GRR result, from 0.55 to 0.71. Removal of the six SDG&E projects resulted in a 26 percent increase in the FY GRR result, from 0.51 to 0.64. Removal of the seven SCG projects resulted in a 46 percent increase in the FY GRR result, from 0.53 to 0.77.

TABLE 4-6: DISCREPANCY FACTORS FOR PROJECTS WITH ZERO OR NEGATIVE GROSS EX-POST MMBTU SAVINGS

ItronID*	First Year Gross Ex-Ante MMBtu	First Year MMBtu GRR	Lifecycle MMBtu GRR	Customer Agreement Date	Discrepancy Factor and Related Change to First Year Gross Ex-Ante MMBtu Savings						
					Inapprop. Baseline	Ineligible Measure	Calculation Method	Inoperable Measure	Operating Conditions	Tracking Discrepancy	Other
E50001	50,981	0.00	0.00	12/16/11	-50,981						
E50002	49,028	0.00	0.00	12/23/13	-49,028						
E55012	28,289	0.00	0.00	02/08/12	-28,289						
E50020	8,660	0.00	0.00	02/06/14			-866		-7,797		
E55033	11,560	0.00	0.00	02/07/11		-11,560					
E55076	6,081	0.00	0.00	05/14/15	-6,081						
E50112	1,886	0.00	0.00	01/27/15				-1,886			
E50599	64	-0.07	-0.07	03/02/14	-42		6		-34		
E55141	2,577	0.00	0.00	02/20/15		-2,577					
F50003	48,073	-0.02	-0.01	04/29/10			-49,164				
F55002	82,769	0.00	0.00	11/13/14		-82,769					
F50005	24,133	0.00	0.00	03/18/13		-24,133					
F55006	14,717	0.00	0.00	02/28/13		-14,717					
F55007	13,680	-0.65	-0.43	01/29/14	-19,014	-3,557					
F55019	4,535	0.00	0.00	01/24/13				-4,535			
F55026	4,387	-0.08	-0.07	05/29/15		-3,809			-923		
F55101	817	0.00	0.00	07/07/14		-817					
G50002	65,520	0.00	0.00	04/25/11		-65,520					
G55008	17,345	0.00	0.00	09/30/14	-17,345						
G55009	16,198	-2.18	-2.90	03/10/14			-51,485				
G50008	8,803	0.00	0.00	06/17/16	-8,803						
G50018	4,328	0.00	0.00	08/22/14	-4,328						
G50029	975	0.00	0.00	03/10/14	-975						
G55107	40	-0.61	-0.41	10/30/14					-63	-1	
H50002	26,364	0.00	0.00	04/07/14		-26,364					
H50004	12,051	-0.01	-0.01	04/11/12				-12,048	-177		
H50005	5,248	0.00	0.00	07/11/13		-3,673					-1,574
H50009	2,543	-0.03	-0.03	11/30/11					-2,628		
H50018	1,292	0.00	0.00	06/12/14	-1,292						
H50027	184	0.00	0.00	12/20/11				-184			
Discrepancy Frequency					11	11	4	4	6	1	1

* None of these projects with zero or negative ex-post savings encountered discrepancies associated with measure counts.

4.4.3 Assessment of Downward and Upward Adjustments to Gross Ex-Ante Savings 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 gross ex-post vs. ex-ante estimates. The fraction of the discrepancy attributable to each factor was also recorded.

Table 4-7 summarizes the downward and upward adjustments by discrepancy factor (including projects with zero and negative gross ex-post MMBtu).[66] This summary includes results for all PAs combined at the top of the table and by individual PA below that. Figure 4-5 displays the same information contained in Table 4-7 for all PAs combined. PA-specific plots can be found in Appendix B.

At the statewide level, downward adjustments affected 150 records and caused a -1.1 MMBtu, or 50 percent reduction, to the 2.3 Million MMBtu gross ex-ante savings estimate for all sampled projects. Likewise, upward adjustments affected 53 records and caused a 0.1 MMBtu, or six percent boost, to gross ex-ante savings. The net reduction for both upward and downward adjustments is negative 1.0 MMBtu, or 44 percent. Aggregate statewide results show that the gross ex-ante savings estimates for the majority of discrepancy factors were more greatly influenced by downward adjustments than upward adjustments. Only the measure counts and tracking discrepancies had slightly net positive impacts.

All factors combined, SCE and SDG&E savings estimates were most affected by downward adjustments, at 54 percent and 64 percent, respectively, versus 50 percent across all PAs. PG&E and SDG&E have smaller, but still significant, downward savings adjustments of 47 and 44 percent, respectively.

At the statewide level, the discrepancy factors that had the greatest influence on adjustments to gross ex-ante savings estimates, both in terms of downward adjustments and net change, were operating conditions, calculation method, inappropriate baseline and ineligible measure. This also held true for SCG, but the strongest drivers varied somewhat across the other PAs. For the other PAs, inoperable measures was among the top four negative drivers. PG&E and SDG&E also saw significant *positive* adjustments due to calculation methods, which resulted in a smaller net reduction (PG&E) or a net positive adjustment (SDG&E) for this discrepancy factor.

At the statewide level inappropriate baselines and ineligible measures had the greatest net (of downward and upward adjustments) effect on gross ex-post MMBtu savings. These factors each led to a 12 percent reduction for downward adjustments alone and both factors also led to a 12 percent reduction on a net

[66] 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.

change basis. These were among the most important factors for all PAs, although PG&E had modest reductions due to ineligibility, and SCE was minimally impacted by inappropriate baselines. The strongest single driver of downward adjustments statewide was calculation methods, although this was somewhat countered by upward adjustments associated with this same discrepancy factor.

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 gross ex-ante savings estimates. The reader is referred to Table 4-7 for the purposes of a more thorough examination of upward effects by discrepancy factor category.

TABLE 4-7: RECORDS WITH EX-POST DOWNWARD AND UPWARD ADJUSTMENTS TO FIRST YEAR GROSS EX-ANTE MMBTU IMPACTS, BY DISCREPANCY FACTOR; STATEWIDE AND BY PA

Discrepancy Factor	n Records with Downward Adjustment	n Records with Upward Adjustment	Gross Ex-Ante MMBtu Savings	Gross Ex-Post MMBtu Downward Adjustments	Gross Ex-Post MMBtu Upward Adjustments	Gross Ex-Post MMBtu Net Change
All PAs						
Calculation Method	58	29		-312,214	114,179	-198,035
Inappropriate Baseline	34	1		-282,050	19	-282,031
Ineligible Measure	21	1		-280,292	33	-280,260
Inoperable Measure	11	0		-62,328	0	-62,328
Measure Count	6	1		-266	1,179	912
Operating Conditions	62	26		-184,407	22,252	-162,155
Tracking Discrepancy	3	6		-150	246	96
Other	7	0		-11,799	0	-11,799
All Records	150	53	2,257,846	-1,133,506	137,907	-995,599
PGE						
Calculation Method	17	12		-81,672	72,003	-9,669
Inappropriate Baseline	14	1		-176,177	19	-176,158
Ineligible Measure	8	0		-49,712	0	-49,712
Inoperable Measure	4	0		-18,688	0	-18,688
Measure Count	0	0		0	0	0
Operating Conditions	13	7		-49,838	5,375	-44,463
Tracking Discrepancy	2	2		-149	85	-64
Other	1	0		-11	0	-11
All PGE Records	40	18	800,066	-376,245	77,482	-298,764
SCE						
Calculation Method	18	8		-153,904	5,089	-148,815
Inappropriate Baseline	4	0		-22,821	0	-22,821
Ineligible Measure	9	1		-135,023	33	-134,991
Inoperable Measure	3	0		-26,557	0	-26,557
Measure Count	0	1		0	1,179	1,179
Operating Conditions	19	6		-30,944	1,730	-29,214
Tracking Discrepancy	0	1		0	0	0
Other	1	0		-5,522	0	-5,522
All SCE Records	44	14	695,184	-374,771	8,031	-366,740

TABLE 4-7: RECORDS WITH EX-POST DOWNWARD AND UPWARD ADJUSTMENTS TO FIRST YEAR GROSS EX-ANTE MMBTU IMPACTS, BY DISCREPANCY FACTOR; STATEWIDE AND BY PA (CONTINUED)

SDGE						
Calculation Method	14	3		-12,821	28,374	15,553
Inappropriate Baseline	5	0		-34,907	0	-34,907
Ineligible Measure	3	0		-30,037	0	-30,037
Inoperable Measure	4	0		-17,084	0	-17,084
Measure Count	5	0		-142	0	-142
Operating Conditions	20	8		-48,045	9,239	-38,806
Tracking Discrepancy	0	3		0	161	161
Other	5	0		-6,265	0	-6,265
All SDGE Records	39	11	231,812	-149,301	37,774	-111,527
SCG						
Calculation Method	9	6		-63,817	8,713	-55,104
Inappropriate Baseline	11	0		-48,145	0	-48,145
Ineligible Measure	1	0		-65,520	0	-65,520
Inoperable Measure	0	0		0	0	0
Measure Count	1	0		-124	0	-124
Operating Conditions	10	5		-55,581	5,907	-49,673
Tracking Discrepancy	1	0		-1	0	-1
Other	0	0		0	0	0
All SCG Records	27	10	530,784	-233,188	14,621	-218,567

FIGURE 4-5: EX-POST UPWARD AND DOWNWARD ADJUSTMENTS TO FIRST YEAR GROSS EX-ANTE MMBTU FOR SAMPLED PROJECTS - ALL PAS

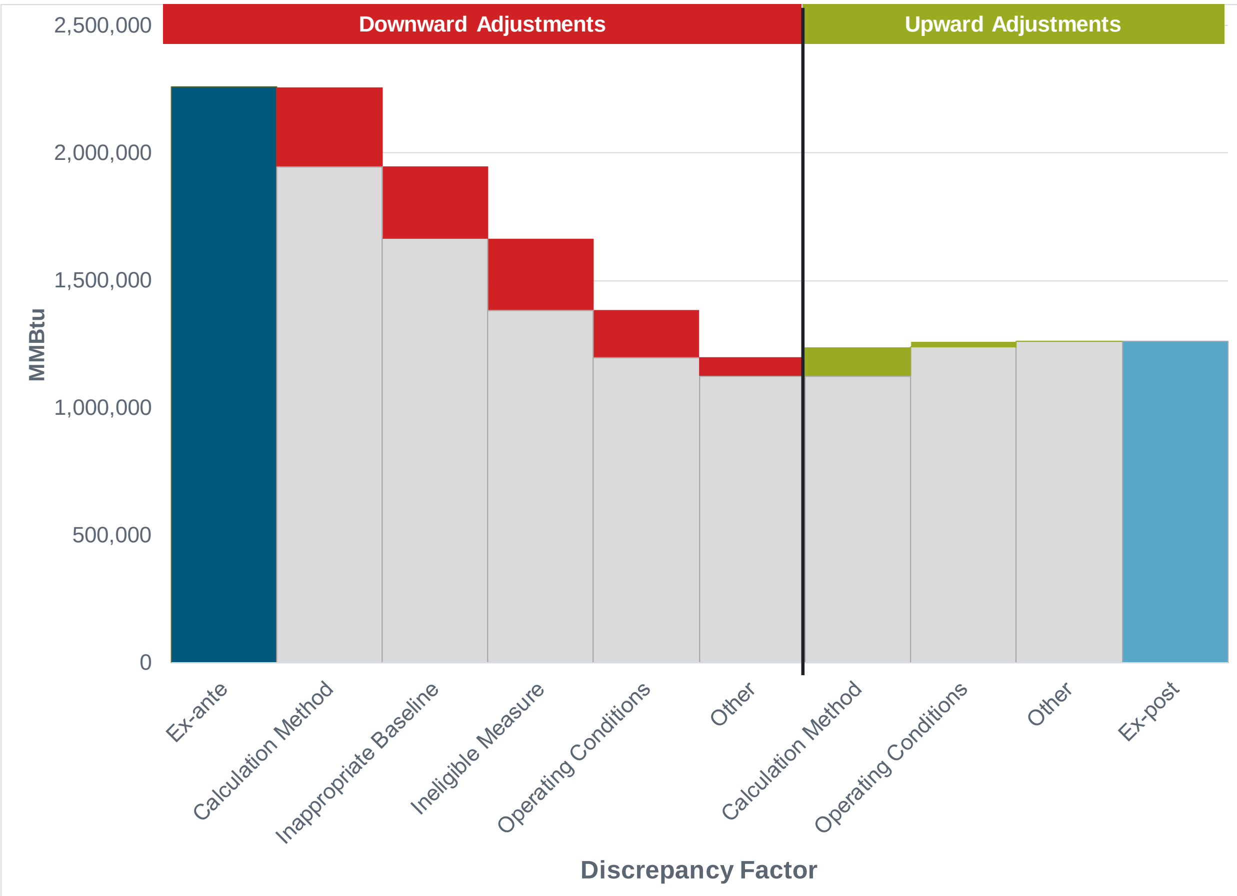
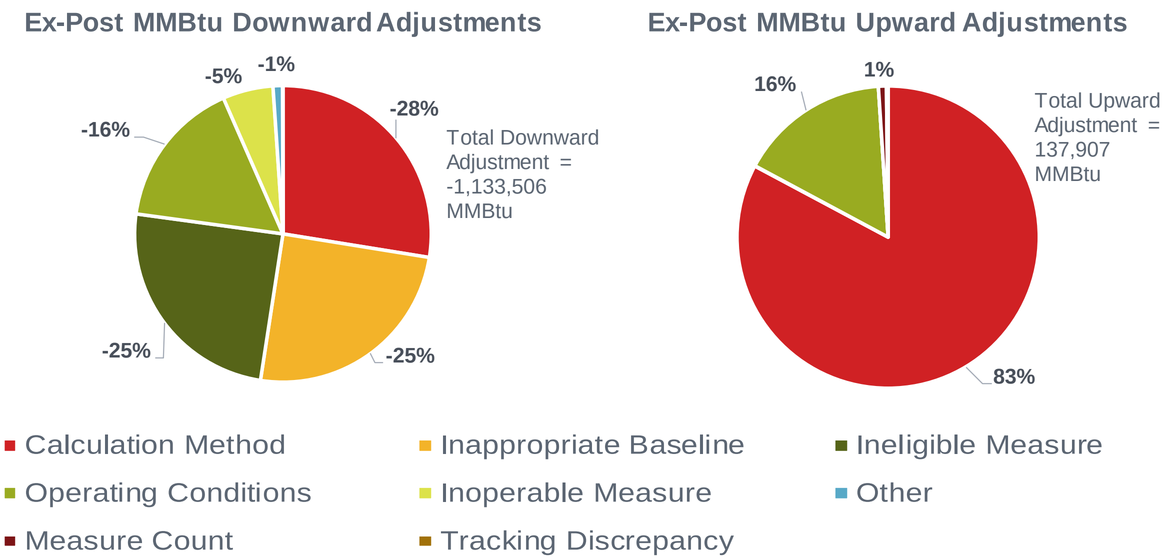


Figure 4-6 shows the distribution of the downward and upward adjustments due to each discrepancy factor across all PAs. Percentages are the fraction of total adjustments (downward and upward, respectively) that are attributed to each discrepancy factor.

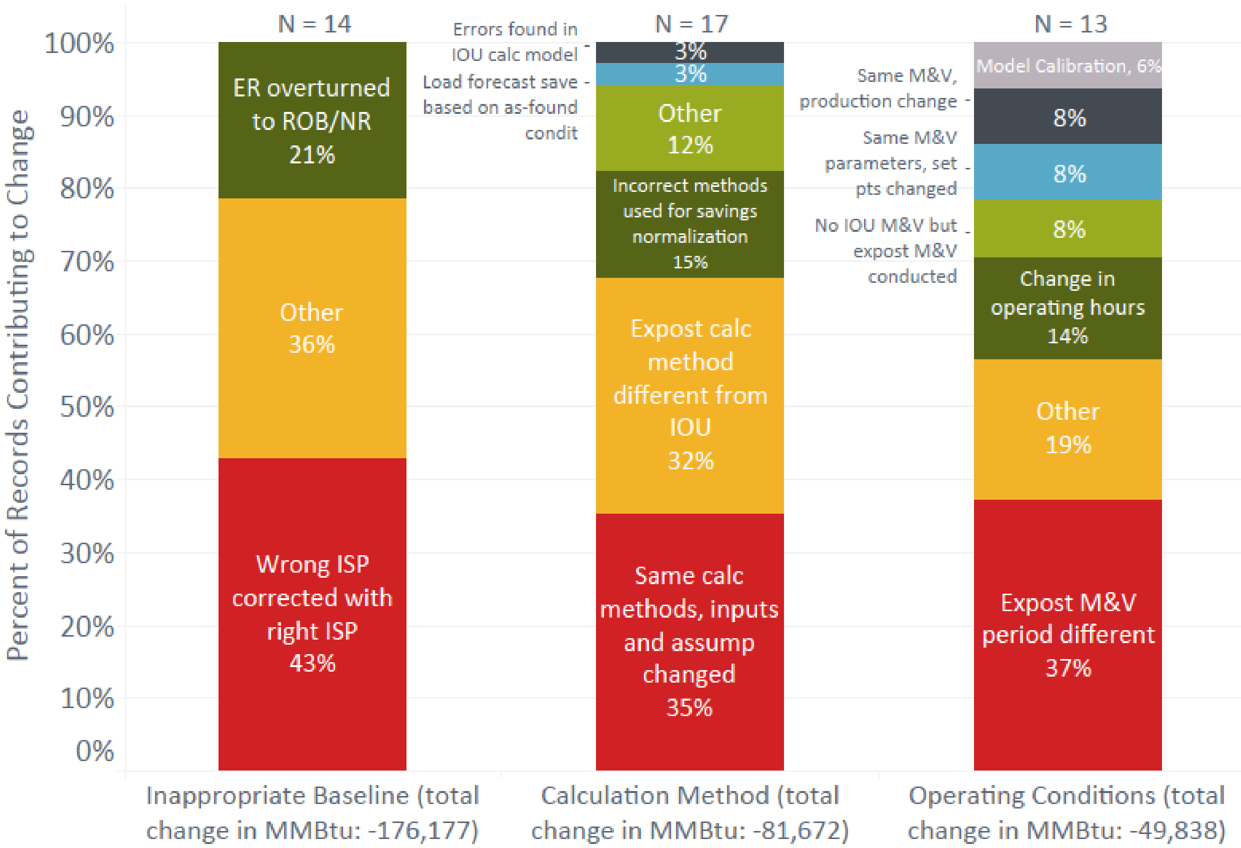
FIGURE 4-6: DISTRIBUTION OF DOWNWARD AND UPWARD ADJUSTMENTS TO FIRST YEAR GROSS EX-ANTE MMBTU BY DISCREPANCY FACTOR - ALL PAS



4.4.4 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, additional explanatory categories were listed with each primary discrepancy factor, and these sub-categories are presented in this section of the report. Figure 4-7 addresses the factors that caused the three largest downward adjustments to gross ex-ante MMBtu for PG&E and provides the frequency (percent of tracking system records) with which each sub-category was noted for each of these primary factors.

FIGURE 4-7: MOST INFLUENTIAL DISCREPANCY FACTORS THAT CAUSED DOWNWARD ADJUSTMENTS FOR PG&E

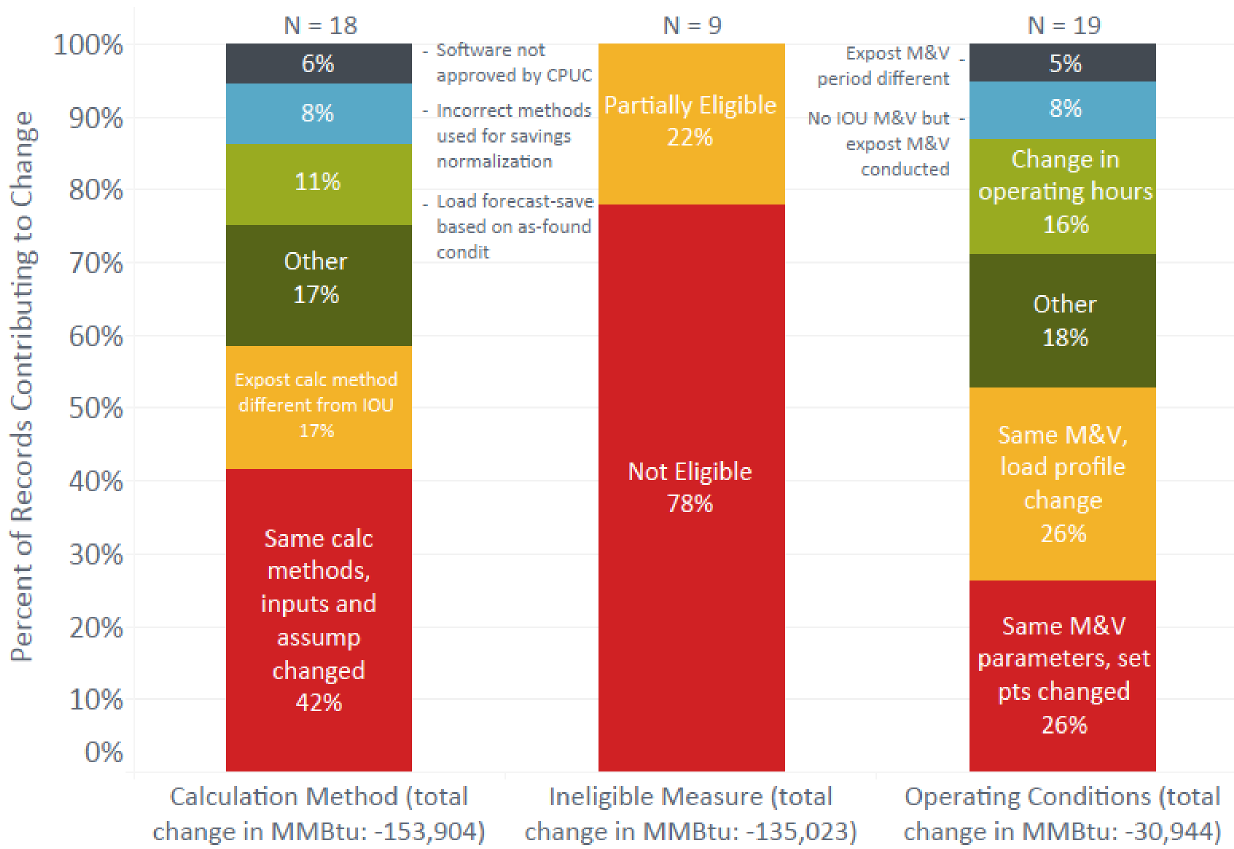


For PG&E the top three discrepancy factors that resulted in a downward adjustment of ex-ante MMBtu impacts were inappropriate baseline, calculation methods, and operating conditions. Inappropriate baseline affected 14 records and resulted in the highest downward adjustment to the gross ex-ante MMBtu savings estimates (-176,177), representing a 22 percent overall reduction to the gross ex-ante MMBtu savings of 800,066. Use of the wrong ISP, corrected with the right ISP by the evaluation was the most frequently observed sub-category within the inappropriate baseline discrepancy group. Other site-specific findings resulted in different baselines, as well as ER being overturned to ROB/NR, representing the second and third most frequent contributing factors to the reduction of gross ex-ante MMBtu savings.

Calculation method changes during the ex-post analysis resulted in the second highest gross ex-ante MMBtu savings reduction and represented a 10 percent reduction to the gross ex-ante MMBtu savings estimate. Seventeen records were affected. The most common sub-categories that explain downward calculation method adjustments were changes to PA inputs and assumptions and gross ex-post calculation methods which were different than IOU methods. The third most commonly observed sub-category involved the use of incorrect methods for savings normalization, followed by a variety of issues that infrequently accounted for downward adjustments to gross ex-ante savings.

Thirteen records had downward savings adjustments due to operating conditions, resulting in a six percent reduction to gross ex-ante MMBtu savings. Over one-third of the downward adjustments were due to ex-post M&V periods which differed from those used for ex-ante savings estimation. The two remaining major categories were made up of "other" issues including changes in IT load and changes in operating hours.

FIGURE 4-8: MOST INFLUENTIAL DISCREPANCY FACTORS THAT CAUSED DOWNWARD ADJUSTMENTS FOR SCE



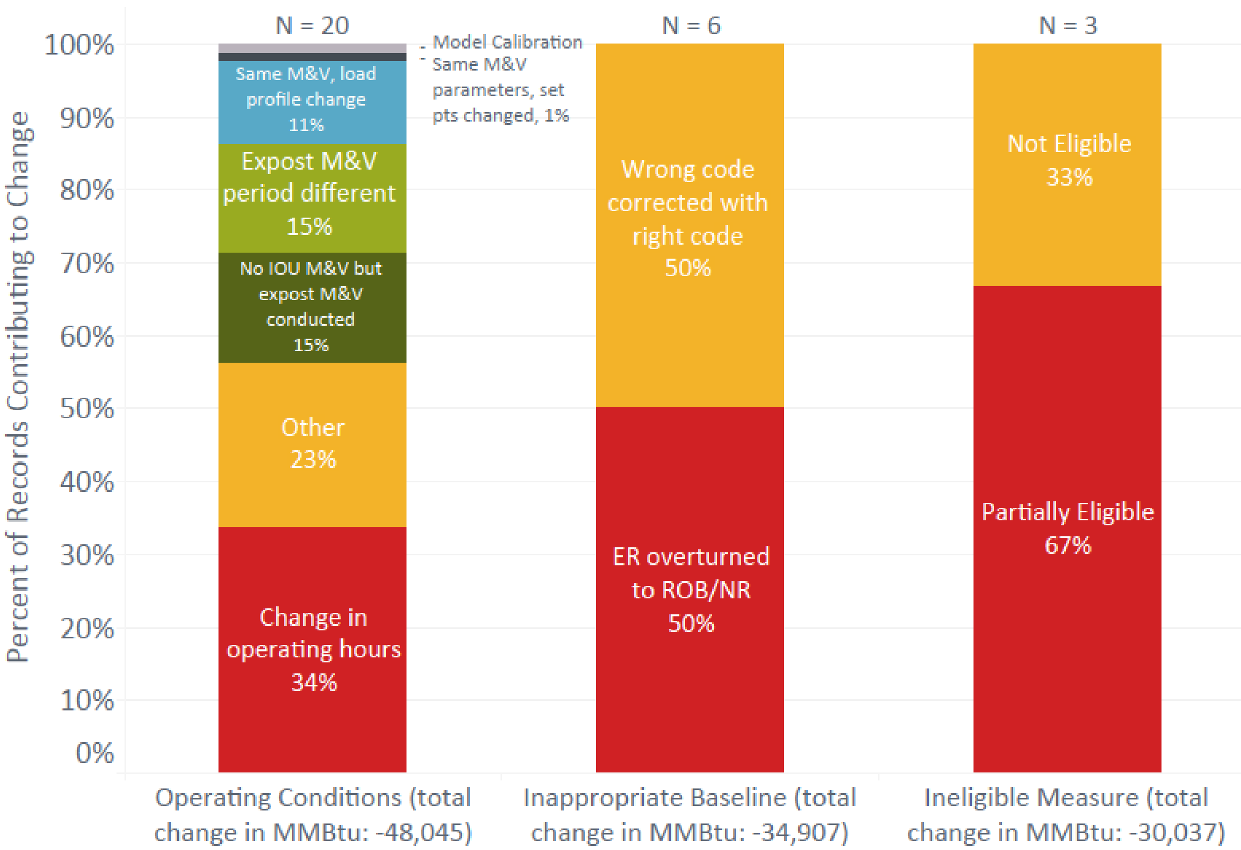
For SCE the top three discrepancy factors that resulted in a downward adjustment of gross ex-ante MMBtu savings were calculation methods, ineligible measures, and operating conditions (Figure 4-8). The eighteen records affected by calculation methods resulted in a 22 percent downward adjustment (-153,903) to gross ex-ante MMBtu savings estimates of 695,184. A range of issues were involved with ex-ante calculation methods, with changes to inputs and assumptions and different calculation methods accounting for almost 60 percent of the issues in this category, with a variety of more minor issues making up the remainder.

Ineligible measures affected 9 records and resulted in the second highest downward adjustment to the gross ex-ante MMBtu savings estimates, representing a 19 percent overall reduction to the gross ex-ante

MMBtu savings. Seven measures were determined to be not eligible, while three measures were found to be only partially eligible. Ineligibility was a result of several reasons, including like-for-like replacements, measures not exceeding baseline, measures being considered routine maintenance, and measures not meeting program requirements.

Changes in operating conditions were found to affect 19 measures, but only resulted in a four percent downward adjustment to the gross ex-ante MMBtu savings estimates. Over 50 percent of these measures were affected by changes to set points and changes to load profiles made by the evaluation team.

FIGURE 4-9: MOST INFLUENTIAL DISCREPANCY FACTORS THAT CAUSED DOWNWARD ADJUSTMENTS FOR SDG&E

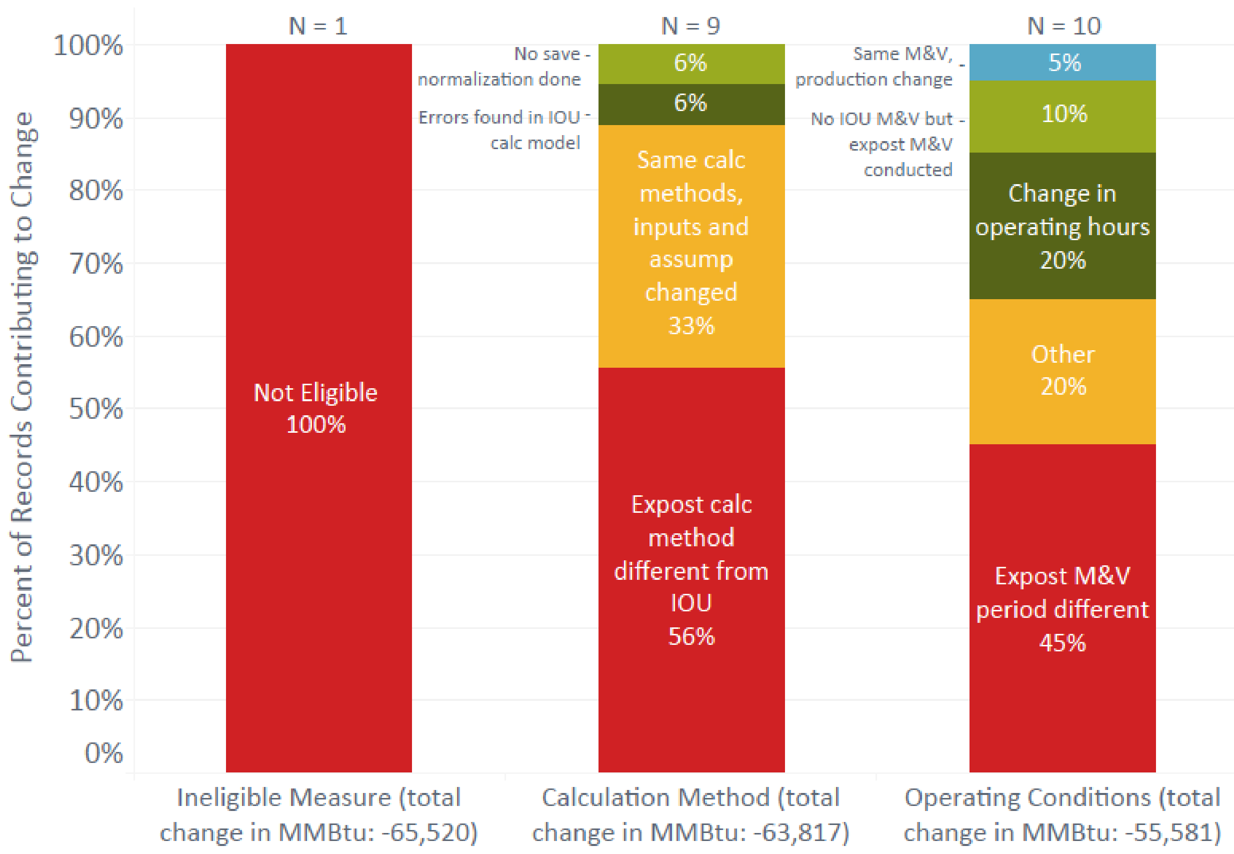


For SDG&E the top three factors that resulted in a downward adjustment of gross ex-ante MMBtu impacts were operating conditions, inappropriate baseline, and ineligible measures (Figure 4-9). For 20 records with downward effects due to operating condition, changes during ex-post analysis resulted in a 21 percent reduction (-48,045) to gross ex-ante MMBtu savings estimates of 231,812. Changes to the operating hours was the most dominant sub-category, affecting 34 percent of the measures in the category, followed by other issues (which include buildings not reaching full operation and equipment not functioning as expected), use of ex-post M&V, use of ex-post M&V which included a different period, load profile changes, changes in set points, and model calibration-related changes.

Inappropriate baselines were the second largest reason for downward adjustment to SDG&E's ex-ante MMBtu savings, affecting 6 records and resulting in a 15 percent decrease in gross ex-ante MMBtu impacts. Sub-categorical explanations were led by ER projects being overturned to ROB or NR, and wrong codes being corrected with the right codes.

Three SDG&E measures were determined to be ineligible (two partially eligible, and one not eligible), causing the third largest reason for reduction to SDG&E's gross ex-ante MMBtu savings, representing a 13 percent downward adjustment. These measures were found to be ineligible, as they did not reflect efficiency upgrades in one case, and in two cases were considered standard practice.

FIGURE 4-10: MOST INFLUENTIAL DISCREPANCY FACTORS THAT CAUSED DOWNWARD ADJUSTMENTS FOR SCG



For SCG the top three reasons that resulted in a downward adjustment of gross ex-ante MMBtu savings were ineligible measures, calculation methods, and operating conditions (Figure 4-10). Only one ineligible measure was identified but led to the highest categorical reduction to gross ex-ante MMBtu savings estimates. This represents a 12 percent overall reduction (-65,520) to the gross ex-ante savings of 530,784 MMBtu. This measure was found to be ineligible as it involved fuel switching but failed the three-prong test.

There were nine measures where the gross ex-ante MMBtu impacts were also reduced by 12 percent due to calculation methods. The gross ex-post estimates frequently applied different calculation methods or different inputs and assumptions, but other issues included errors in the original calculations and lack of appropriate use of normalization.

Operating conditions represent the third most important factor that reduced SCG's gross ex-ante MMBtu savings estimates. Ten records were affected by this discrepancy factor. The most frequently cited sub-categorical explanation comes from gross ex-post M&V periods which differed, followed by other issues including changes to inputs and non-operational equipment. Changes in operating hours were responsible for 20 percent of record-based differences, and minor changes resulting from a lack of PA M&V and changes to production.

4.5 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. The suggestions reported include some additional steps the PAs and implementers can take to improve ex-ante savings estimates, based on savings gaps that were identified and documented in this evaluation. Those suggestions and considerations are summarized below for each of the main discrepancy factors noted above in Section 4.4 – calculation methods, inappropriate baseline, ineligible measures, and operating conditions. 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[67] where relevant suggestions apply are listed in parentheses.

[67] 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.

4.5.1 Calculation Methods

Changes in calculation methods represents the greatest cause for evaluation-based reduction to gross ex-ante saving estimates.

Ex-Post Calculation Method Different from PA

- Ex-ante regression models should be informed by longer duration trend data. Furthermore, for ex-ante regression models involving both energy consumption data and production data (i.e., energy intensity), the PAs should attempt a variety of models using differing time intervals, such as daily versus hourly, in order to identify model-based estimates with the best fit regression curve (F50003).
- Where regression models are used, make sure that the R squared values are 0.70 or higher and the CV(RMSE) values are lower than 15 to 20 percent (H50003).
- For pump efficiency improvement projects, historical energy usage and production data should be used to derive estimates of kWh/acre-foot and OPE. The ex-ante calculations were based on pre- and post-retrofit pump tests, in order to set OPE values, which are instantaneous efficiency estimates and do not normally represent longer-term pumping operations (E50042).
- Calculated savings should be based on robust data sets representing longer-term and stable operation of equipment and systems. PAs should collect appropriate trend data that demonstrate typical operation, and ensure that M&V data used to estimate ex-ante savings properly account for variation in weather, seasonality, equipment performance and production schedules/operations. Where variability is present, PAs should wait to claim savings until a more confident savings estimate, based on typical operation, has been developed (E55126, E55002, F50001).
- PAs should use proven tools for estimating savings (E55268). For this project a 3rd party tool was used to estimate savings, but the resulting savings estimate differed substantially from accepted work paper-based savings estimates.
- PAs should ensure that modeling and related inputs and assumption are consistent with observed system operation (F50006, E50014, F50019, G50009, G55015). This includes use of pre- and post-installation M&V, including measurement of calculation parameters and collection of relevant production records, and use of measured versus assumed inputs.
- Follow the guidance document posted by the CPUC on estimating savings when non-PA supplied energy sources are used (Energy Efficiency Savings Eligibility at Sites with non-IUO Supplied Energy

Sources – Guidance Document, Version 1.1).[68] For example, performing an hourly net grid impact analysis if onsite generation is significant (F50019, H50026, F55024).

- For certain applications, such as where the baseline is represented by the pre-existing equipment and pre- to post-installation conditions are stable and well documented, use of an IPMVP Option B or C regression model may be preferable to other calculation-based approaches. Regression models should also account for all non-routine adjustments, as facilities often undergo changes unrelated to program efficiency-based improvements, and savings estimates should be normalized for production and weather differences. It is also critical that the measure-impacted accounts be properly identified and used in regression models. Regressions may serve to better bound the savings and may also be used as a sanity check of results derived using other calculation approaches (F55001, E55126, F55080).
- The selected PA savings calculation approach should be informed through collection of M&V data, especially post-installation trend data where feasible and warranted (E55001).
- The PAs should ensure that projects have an identifiable and documented case for energy efficiency claims. This should include the validation of equipment specifications and performance used in the calculations and some level of post-installation M&V that assures savings claims, prior to final project approval (G55009, E50003).
- Peak demand savings estimates should be based upon the DEER-defined peak demand reduction period for a given climate zone (E55126, F50803, H55901).
- PAs should encourage participating customers to collect and retain data for purposes of conducting project-level M&V, especially where instrumentation is available. This will be useful for both evaluators and customers wishing to conduct self-review/verification of project-level savings; this is especially needed for larger projects (E55004).
- Perform a sanity check using simple rules-of-thumb in order to be confident in calculated results (H50021). For example for a pump efficiency improvement project, compare dedicated electric usage against modeled usage, given facility production levels. For this project the ex-ante savings were based on a motor running at 100 percent speed, but ex-post estimates suggest that the motor normally runs at roughly 60 percent speed.
- For NRNC whole-building projects the PAs should use the non-compliance mode to estimate savings and compliance mode to demonstrate project eligibility (H55903, H50802, H55908, E50801).

[68] http://www.cpuc.ca.gov/WorkArea/DownloadAsset.aspx?id=11610

Errors Found in PA Calculation Model

- PA savings calculations should accurately reflect the system conditions being modeled (E50006).
- Institute installation report review of calculations and input parameters to ensure that errors are identified and corrected (F55016, E50020).
- PAs should not develop impact calculations that are based on mixed treatment of baseline and project equipment (F50004). For example, modeling of post-installation chiller performance curves using metered data, but basing baseline equipment performance on a code-based performance curve.

Same Calculation Methods, Inputs and Assumptions Changed

- PAs should use a mix of measured data and, secondarily, manufacturer equipment specifications to inform calculation inputs (F50259, F50383, G50011, G55013, E50624).
  - For one project ex-post spot kW readings of the pump demonstrated a lower loading profile than was assumed in the ex-ante calculations of savings.
  - Another project included ex-ante inputs and assumptions that were unsubstantiated.
  - For another project the ex-post evaluation found that the actual wastewater flow rate was just 17 percent of the assumed ex-ante value.
  - For another project flue gas analysis was used to measure combustion efficiency of the boiler, which was vastly different than the rated combustion efficiency used in the ex-ante savings calculations.
  - For another project the installed pump efficiency was overstated at 95.8 percent and corrected in the ex-post calculation to 95 percent based on nameplate data.
- Where the PAs specify existing equipment as the baseline, the existing equipment must be capable of meeting post-installation production levels (E55043). For this project, involving pump efficiency improvements, the ex-post estimates were capped at a production level that the existing pump was capable of meeting.
- For projects entailing new processes, where pre-existing equipment conditions cannot inform ex-ante savings estimates, post-installation M&V should be used to inform inputs and assumptions, including sparing use of conservative assumptions where collection of measured data is infeasible or impractical for a given parameter (G50025).
- Where models are informed by data trends, the derivation of analysis inputs should carefully account for periods of unstable operation and varying performance; ultimately the inputs should represent typical performance (E50010).

- PAs should carefully document all inputs to models, especially where data are used to derive model inputs, such that evaluators can reproduce or validate model inputs (F50004).
- PAs should ensure that the final model used to derive savings claims is retained in project files and made available to evaluators when requested (F55024). For this project the ex-post evaluation adjusted model inputs to reflect trend data and updated the weather file.
- PA models should use custom rather than deemed variables in calculations where inconsistencies exist between the project conditions and the deemed approach, such as differences in building type or systems (E55047). Furthermore, for this project control sequences should have been based on the actual control approach rather than an assumed control approach.

Incorrect Methods Used for Saving Normalization

- Production normalization should be performed to account for varying production (E50042, H55015, G50011). One such project involved pump production levels, another involved CFM of compressed air, and the third instance required normalization of water usage.

4.5.2 Inappropriate Baseline

Inappropriate use of baseline represents the second greatest cause for evaluation-based reduction to ex-ante saving estimates.

Wrong ISP/code Corrected with Right ISP/code

- PAs should demonstrate the availability of selected baseline equipment when establishing ISP (E50001). Ordinarily this would include obtaining quotes for available new, less efficient, but functionally equivalent equipment (baseline). 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. For this project the PA selected used/refurbished equipment for the purposes of setting the baseline for a new construction project; use of a used/refurbished equipment baseline is not allowable or appropriate.
- PAs should demonstrate that baseline equipment selected represents a feasible option, given facility constraints and production needs (E55012). For this project the program-installed process cooling system was determined to be the only feasible solution to meet expanded production capacity, and the ex-post evaluators rejected the ex-ante baseline consisting of the same type of process cooling equipment already present at the site, due to physical space limitations, electric service constraints, and production needs.
- PAs should specify baseline equipment that are functionally equivalent to the installed measure (E55076, E50599).

- For one project the new production equipment was being installed in order to meet expanded market demand for a different product, and the installed equipment was found to be the only viable equipment choice.
- For another project the ex-post baseline pump impeller size for a newly installed booster pump was adjusted to meet design loads, and a corrected pump curve was applied.

- PAs should ensure appropriate application of code when establishing baseline (H55001). For this project lighting equipment baseline LPD values were inconsistent with code requirements in the ex-ante estimates.
- Where existing conditions and associated equipment efficiency levels are governed by program requirements for previously installed equipment, claimed savings are only allowable that exceed previous program requirements (G50008). For this project the ex-post savings estimates were associated with heat recovery from on-site generation equipment that was originally procured with financial support from the Self Generation Incentive Program (SGIP), which at the time required a prescribed level of heat recovery. Since this project involves re-instituting heat recovery, the SGIP heat recovery requirements constitute the baseline.
- PAs should rely on previous EAR and ex-post evaluation conclusions for the purposes of setting project baselines (G55008). For this project the installed measured had previously been determined to be the baseline in industry leading nut drying facilities.
- PAs should conduct independent research for the purposes of identifying project-level ISP baseline (in accordance with the Industry Standard Practice Guide, Version 1.2A)[69] and provide a comprehensive narrative backed up by data that correctly identifies ISP (E50014). The ex-ante calculations for this project suggested that the existing furnace was ISP, but the evaluation conducted independent ISP research that concluded this is not the case, and that ISP is best represented by higher efficiency models available in the market.
- When using simulation models, demonstrate that baseline systems meet code-based energy budget allowances by running performance compliance (E50245). For this project it was found that the ex-ante modeled baseline HVAC unit was not capable of meeting Title 24 requirements, and the ex-post baseline was upgraded to a VAV system from a two-deck, multizone system.
- For new oil production steam generator applications, consultation with industry experts found that split pass design is 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 (E55004, E55002). Furthermore, for these two projects the evaluators set the baseline steam generator efficiency to 82 percent, as directed by CPUC staff.

[69] http://www.cpuc.ca.gov/WorkArea/DownloadAsset.aspx?id=5315

- For NRNC and other 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 (E50801, H55903). This consists of checking the baseline INP file and making sure it is consistent with the ACM manual, and if not, then modify the model and run the simulation using a DOE-2.1E engine.
- During application review the PAs should carefully consider all relevant code requirements and update ISP and other baseline determinations for relevant measures (G50029, E50624).
  - For one project Title 24 was found to require the installation of a pool cover, thereby overturning the ex-ante baseline of no pool cover.
  - For another project ISP was found to be a 94.8 percent efficient motor versus an ex-ante modeled baseline of 94.4 percent; careful research should be conducted for non-standard equipment such as hollow shaft motors, using online searches and calls with vendors.

ER Overturned to ROB/NR

- For ER claims the condition and functionality of the existing equipment should be documented, along with the RUL, and it should be demonstrated that the existing equipment are able to meet post-installation loads and functional requirements. Preponderance of evidence that the program induced the early retirement should be used to establish early retirement claims (F55039, H55002, E55047, H50007, E50002, G50021, G55033).
  - For one project the replaced EMS system was more than 10 years old and not working properly; the evaluation overturned the project type to natural replacement, thereby rejecting the ex-ante existing equipment baseline and early retirement claim.
  - For another project increased cooling loads in the post-installation condition led the ex-post evaluators to conclude that the existing equipment were unable to meet the increased loads, and therefore rejected the ER claim in favor of a natural replacement project designation and a code-based baseline chiller efficiency level.
  - For another project the PA failed to adequately document ER claims using preponderance of evidence.
  - For another project a steam boiler was replaced with a hot water boiler, and because the existing piping also had to be replaced, this project is considered to be a major renovation/natural replacement project with a hot water boiler baseline. Furthermore, the installed boiler controls do not exceed code requirements or standard practice, and associated savings were disallowed; the decision to reject ER claims was amplified by the condition of the existing pneumatic controls, which were not functioning properly and required frequent maintenance.

Other

- The PAs should carefully investigate and document the age, condition and functionality of existing equipment and operations, and use these to establish proper baselines (F55007). For one project the PA selected a hypothetical, minimum efficiency, electric equipment option, but the evaluation overturned this baseline based on the fact that the existing methane/biogas equipment were still in good working order, with significant remaining useful life, and the PA had not adequately demonstrated that an electric equipment baseline was a reasonable assumption to make.
- PAs should ensure that baseline and efficient project equipment operational parameters are properly documented and modeled (E50037). For this project M&V data were used to negate claims of reduced equipment cycling and associated natural gas savings claims.
- PAs should also document and ensure the accuracy of baseline operational assumptions (H55030). For this project the PA claimed outdoor lighting measure impacts during the summer coincident peak, which occurs during the daytime when outdoor lights are turned off.
- When baseline conditions are defined by the pre-existing system, the PAs should utilize measured data to define those conditions, select a representative baseline period, and thoroughly document the pre-existing conditions for the purposes of establishing baseline (G55003). For this project the ex-post analysis selected alternative billing data time periods that better represent baseline.

Regressive Baseline

- PAs should push their customers into incremental energy savings over non-regressive baseline equipment (G50018). For this project it was discovered that a new ozone generator replaced an old ozone generator that was not functioning well, resulting in like-for-like equipment replacements and zero ex-post savings.
- 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 (H50018, G50013).
  - For one project laminar flow restrictors were installed on faucets, but it was discovered that the removed faucets included laminar flow restrictors, thereby leading to no ex-post savings, in conformance with non-regressive baseline rules.
  - For another project the ex-ante savings were based on an assumed baseline of no heat curtain, but it was discovered that a double layer heat curtain was the pre-existing condition; verification and vetting of baseline claims should take place at various stages of the application process, especially savings calculation review.

4.5.3 Ineligible Measures

Ineligible measures represent the third greatest cause for evaluation-based reduction to ex-ante saving estimates.

- PAs should confirm that projects submitted exceed code requirements that were in place at the time a given facility was built and/or based at the time of subsequent “alterations” to project-affected equipment and systems (E55033, E55141, F55026).
- PAs should review the natural gas tariff for participating customers to ensure that PPP charges apply (E55019). Projects associated with gas use reduction are ineligible if PPP charges are not associated with the gas account where savings are realized.
- Fuel switching projects must pass the 3-prong test (G50002) to be eligible.
- Maintenance measures are ineligible as custom projects (F55002).
- PAs should use preponderance of evidence to support early retirement claims (E55047). For this project ER was overturned to ROB which in-turn led to three out of six measures becoming ineligible due to mandatory Title 24 requirements.
- PAs should confirm that projects meet program eligibility performance thresholds, for example, the Savings by Design time dependent valuation (TDV) savings margin (F55101).
- 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 (H50002, F50005. F55006).
- When 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 (E55019).
- 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 need to assess the minimal technical requirements of a given project that meets all service requirements. This also includes a careful assessment of the condition of the pre-existing equipment for the purposes of selecting an appropriate project type and associated baseline. Finally, this includes a review of prior CPUC evaluation findings, EAR directives and guidance, seeking guidance from the CPUC during project development and review, and conducting ISP research where warranted (E50037, F55007, F55005, E50016, E50009, H50005).

4.5.4 Operating Conditions

Changes in operating conditions represent the fourth greatest cause for evaluation-based reduction to ex-ante saving estimates.

M&V Improvement Opportunities

- True-up savings based upon post-installation data, including the use of observed model inputs and conditions (H50009, E55006, E50020, H55004, F55024, E50801, F50069, E55100, F50355, H55013, F55012, H50801, H55908, F55026, G50005, H50802, F55906, H55903, F50029).
  - Verify that savings calculations are based on actual occupancy schedules and reflect the post-installation conditions accurately. Identify any changes to system operating conditions -- door openings, laundry washed per year, economizer enthalpy type, chilled water setpoint, heating hot water supply setpoint, temperature setpoints, equipment operating schedules, solar fraction, control strategies, fan control settings, fan speed, chiller efficiency (based on trend data), cooling tower load shapes, boiler efficiency (based on flue gas analysis records), economizer high limit temperature and DAT reset range, SAT reset, minimum outside air settings, and VFD speeds. Savings models should be adjusted to the new operating conditions after ensuring that measure operation and production levels are stable.
  - 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 more thoroughly document input sources used to develop ex-ante savings estimates (G55032, G55015, E55185, E55185, F55080).
  - In one instance the rated efficiency of the boiler was adjusted based on equipment specifications and the CFM capacity of the air handlers were adjusted to reflect air balance test results.
  - In another instance the ex-post evaluation used natural gas trend data and observed conditions to update ex-ante assumptions.
  - In two instances PA documentation did not support peak hour savings claims.
  - In one additional instance interval data demonstrated that the equipment do not operate during the DEER-defined peak demand period.
- 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 during post-installation project reviews or M&V (F50012, H55033, F50259, H55013). For these particular projects the ex-post evaluation found conditions that varied substantially from

ex-ante inputs and assumptions, including system pressure drops, pump loading, number of cars washed, and amount of laundry washed.

- PAs should use trend data to generate performance curves and estimate power consumption instead of using default curves (F50016, E55014). Parameters where this was feasible for these two projects includes pump power and performance, chiller performance, cooling tower fan power, AHU power, and CRAH fan power.
- 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 (E55080, F50004, G55004, G55010, G50001, H50001, E50006, H50003).

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. Conduct due diligence to ascertain that the annual operating profile of equipment is based on representative data, especially for variable loads, including seasonal variation in production (F55154, F50029, H55030, F50030).
- 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 (G55059).
- In order to claim kW savings, verify that equipment operates during the coincident peak period using the DEER definition of peak (E55001).

Measure-Specific Issues

- For agricultural pumping projects operating conditions might need to be based on a longer duration of pre- and post-installation operation than a single year (E50599, G50003). Operations can be affected by weather/drought conditions and water availability.
  - In one instance fields were fallowed for 2 years, substantially reducing pumping operations
  - In the other instance water restrictions were in effect resulting in a drop in production.
- Ensure that equipment are operating in accordance with the characteristics that lead to claimed savings (H55007, H50004, G55107). For example, if pump or fan modulation is associated with project savings claims. Equipment operation should be verified in accordance with modeled savings estimates, such as validating that heat reclamation systems are operational.

- For new construction projects associated with either tenant improvements or new buildings, wait to file claims once the project is fully built out and occupied. A certificate of occupancy can be used to inform the timing of claims. CPUC evaluation guidance is to model savings based on the as-found conditions (H55030, H55011, H55011).
- For deemed-like measures – for example, simple lighting measures and small HVAC projects – the PAs should apply DEER methods or direct claims through deemed measure program channels (H55028, H55029).

5 NTG RESULTS

The methodology used to develop individual, site-specific net-to-gross (NTG) estimates is summarized in the IALC Research Plan.[70] 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.[71]

5.1 NUMBER OF COMPLETED SURVEYS

For the IALC 2015 study, a total of 208 NTG surveys were completed. 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.[72] In total, 115 of the completed NTG sample points overlapped with the 148 evaluated gross M&V points. Table 5-1 below reports the number of completed telephone surveys by utility, including the number of main versus backup points used, and the percent of first year gross ex-ante MMBtu claims represented. Each utility accounted for roughly one-quarter of the completed surveys, with SCG having the largest number of completes (56) and SDG&E having the least (44). The surveys completed represent 39 percent of FY gross ex-ante MMBtu savings.

70 http://www.energydataweb.com/cpucFiles/pdaDocs/1198/IALC_Research_Plan_Final_11-12-2014.pdf
http://www.energydataweb.com/cpucFiles/pdaDocs/1541/2015%20Custom%20Research%20Plan%20Addendum.docx

71 The IALC Custom NTG surveys also include a battery of questions to address participant spillover. However, estimation of spillover is not part of the IALC scope of work. For 2013-2014, these data were 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

72 Backup completes typically indicate that a complete with one of the main points was not possible. This is due to a combination of factors, including non-response, project contacts that are no longer employees of the target facility, contact not successful after multiple attempts, disconnected phone and bad contact information. Note that robust NTG results are dependent on a statistically representative sample in each sampling domain and are not dependent on main vs backup points.

TABLE 5-1: COMPLETED SURVEYS BY PROGRAM ADMINISTRATOR

Program Administrator	Completed NTG Points (n)		Percent of FY Gross Ex-Ante MMBtu Savings
	Main	Backup	NTG Sample
PG&E	39	15	29%
SCE	40	14	42%
SDG&E	31	13	64%
SCG	27	29	66%
All PAs	**208**		**39%**

5.2 WEIGHTED NTG RESULTS

Weighted results are presented in this section for each sampling domain (PA and project size strata). 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 gross 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 5-3, Table 5-4, Table 5-5 and Table 5-6 below are based on the removal of 13 surveyed projects, leaving a total of 208 sample points. This was due to either an ineligible measure (eight projects removed) or inappropriate baselines (five projects removed). In general, these 13 projects (Table 5-2) 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 5-2: SURVEYED PROJECTS REMOVED FROM NTG ANALYSIS

ItronID	Reason for Removal
E50001	The measure was found to have an inappropriate baseline resulting in zero savings.
E55012	The measure was found to have an inappropriate baseline resulting in zero savings.
E55033	The measure was found to be ineligible, therefore, there is no legitimate project to evaluate.
E55141	The measure was found to be ineligible, therefore, there is no legitimate project to evaluate.
F55002	The measure was found to be ineligible, therefore, there is no legitimate project to evaluate.
F50005	The measure was found to be ineligible, therefore, there is no legitimate project to evaluate.
F55006	The measure was found to be ineligible, therefore, there is no legitimate project to evaluate.
G50002	The measure was found to be ineligible, therefore, there is no legitimate project to evaluate.
G55008	The measure was found to have an inappropriate baseline resulting in zero savings.
G50029	The measure was found to have an inappropriate baseline resulting in zero savings.
H50002	The measure was found to be ineligible, therefore, there is no legitimate project to evaluate.
H50005	The measure was found to be partially ineligible, therefore, there is no legitimate project to evaluate for electric savings (gas savings were non-zero, and NTG results were applied on gas savings only)
H50018	The measure was found to have an inappropriate baseline resulting in zero savings.

5.2.1 PG&E Combined Electric and Gas

Table 5-3 below reports NTGR results for PG&E. The resulting weighted average NTGR of 0.53 for 2015 is similar to the 2013 and 2014 NTGR values of 0.55 and 0.51, respectively.

TABLE 5-3: WEIGHTED NET-TO-GROSS RATIOS FOR PG&E – COMBINED ELECTRIC AND GAS

Sampling Strata	MMBtu NTGR	N Sample Frame	n NTGR Sample	Percent of FY Gross MMBtu Sampled
1	-	0	0	-
2	0.53	17	11	63%
3	0.59	25	12	48%
4	0.51	105	16	15%
5	0.53	1,204	15	2%
All - Weighted NTGR	**0.53**	**1,351**	**54**	**29%**

90 Percent CI	0.49 to 0.58
Relative Precision	0.08
n NTGR Completes	54
N Pop Sampling Units	1351
ER	0.35

Results were very consistent across sample size strata, with all stratum-level scores in the 0.51 to 0.59 range. Highlights include the following:

- Oil and gas company projects had average NTGR results ranging from 0.38 to 0.70. The majority of project NTGRs were in the 0.50 range, and represented a single technology, a split pass steam generator with VFDs.
- Water/wastewater projects continued to exhibit low NTGRs, in the 0.00 to 0.53 range.
- Datacenter project NTGRs were moderate. Most fell into the 0.50 to 0.60 range, although two smaller projects had NTGRs of just 0.36. In general, standalone / dedicated datacenter facilities are highly motivated by competitive pressures to reduce their operating costs, of which electricity costs are a significant component. In light of this, the program needs to carefully examine its role in these decisions and assess how to reposition itself to influence customer decisions on efficiency above and beyond what is already taking place.

5.2.2 SCE Electric

Table 5-4 presents SCE NTGR results. The resulting weighted average 2015 NTGR is 0.57, which is the same result seen in 2013 and an improvement relative to 2014 (0.46).

TABLE 5-4: WEIGHTED NET-TO-GROSS RATIOS FOR SCE – ELECTRIC

Sampling Strata	MMBtu NTGR	N Sample Frame	n NTGR Sample	Percent of FY Gross MMBtu Sampled
1	0.71	6	5	81%
2	0.56	14	8	58%
3	0.51	33	12	39%
4	0.53	87	13	16%
5	0.50	625	16	4%
All - Weighted NTGR	**0.57**	**765**	**54**	**42%**

90 Percent CI	0.53 to 0.62
Relative Precision	0.07
n NTGR Completes	54
N Pop Sampling Units	765
ER	0.34

Average NTGR results across all strata ranged from 0.50 to 0.71.

- Stratum 1 project NTGRs were generally good, with NTGRs ranging from 0.53 to 1.0. The three largest projects (of five projects in this size stratum) had NTGRs of 0.70 and higher, reflecting high program influence for these projects. This stratum is comprised of a mix of business types including chemical, metal and mineral production facilities and one research facility.
- Project-level NTGRs in the remaining strata varied widely, and ranged from 0.0 to 0.95. These strata contain a blend of projects with high, medium and low NTGRs. These results suggest that a mix of motivations is present in decision making, involving both program considerations and non-program factors.
- Projects undertaken by municipal water agencies generally had low NTGRs, in the range of 0.20 to 0.57. Municipal water/wastewater projects were flagged in previous Custom program evaluations as being prone to high free ridership.
- Datacenter project NTGRs again were moderate, with average values around 0.50.

5.2.3 SDG&E Combined Electric and Gas

Table 5-5 presents 2015 NTGR results for SDG&E. The weighted average 2015 NTGR for SDG&E's electric and gas projects is 0.50, which is approximately the same result as 2014 (0.51). These scores represent a decline from the SDG&E 2013 NTGR result of 0.59.

TABLE 5-5: WEIGHTED NET-TO-GROSS RATIOS FOR SDG&E – COMBINED ELECTRIC AND GAS

Sampling Strata	MMBtu NTGR	N Sample Frame	n NTGR Sample	Percent of FY Gross MMBtu Sampled
1	0.44	1	1	100%
2		3	2	61%
3	0.61	14	10	73%
4	0.49	23	10	46%
5	0.34	57	21	30%
All - Weighted NTGR	**0.50**	**98**	**44**	**64%**

90 Percent CI	0.44 to 0.57
Relative Precision	0.13
n NTGR Completes	44
N Pop Sampling Units	98
ER	0.68

Results varied somewhat across sample size strata ranging from a low of 0.34 (stratum 5) to a high of 0.61 (stratum 3). Strata 1 and 2 were combined due to a low number of projects in the sample frame for these strata.

- Public and private university projects continued to perform strongly. In stratum 1/2, there was one strong performing university project (0.73) that accounted for 53% of the sampling weight, and Stratum 3 had the highest NTGR value for SDG&E, which again was driven by three high performing university projects with NTGR values between 0.73 and 1.0.
- There were a number of poorly performing water/wastewater projects in Stratum 5 with NTGRs of between 0.20 and 0.25. These pulled the average NTGR down for Stratum 5 projects in general, and it resulted in the lowest NTG score for any SDG&E strata.
- Again, datacenter projects had low to moderate NTGRs, spanning from 0.18 to 0.59.

5.2.4 SCG Gas

For SCG gas projects, the weighted average NTGR result is 0.57, as shown in Table 5-6 below. This result is down somewhat from the 2013 and 2014 NTGR values of 0.66 and 0.62.

TABLE 5-6: WEIGHTED NET-TO-GROSS RATIOS FOR SCG – GAS

Sampling Strata	MMBtu NTGR	N Sample Frame	n NTGR Sample	Percent of FY Gross MMBtu Sampled
1	-	0	0	-
2	0.61	7	6	82%
3	0.64	7	5	67%
4	0.46	20	11	56%
5	0.49	162	34	27%
All - Weighted NTGR	**0.57**	**196**	**56**	**66%**

90 Percent CI	0.54 to 0.61
Relative Precision	0.07
n NTGR Completes	56
N Pop Sampling Units	196
ER	0.37

NTGR results by size stratum ranged from a low of 0.46 for Stratum 4 to a high of 0.64 for Stratum 3. Note that there were none of the largest Stratum 1 projects in the population. In general, smaller size projects, those in Strata 4 and 5, had lower NTGRs.

- Oil and gas company projects generally had high NTGRs, with values of 0.75 and higher. These strong projects are primarily focused on energy savings and the program incentives help reduce up-front costs and meet internal rate of return requirements. The one exception was a large project from an oil refining company in Stratum 2, which carried a stratum-weight of 26 percent and had a NTGR value of 0.38.
- Projects at public universities and colleges also exhibited high program influence, generally with NTGR values of 0.75 and above. Funding availability for these projects is limited, and the program incentive reduces the amount of up-front expenses. This allows these customers to leverage their limited funding for projects which would otherwise not be viable.

5.2.5 Comparison of 2010-12, 2013, 2014, and 2015 NTG Results by PA – Combined MMBtu

An analysis of NTGR trends since the 2010-12 program cycle reveals that results have shown little improvement since 2010-12. In the 2010-12 evaluation, sampling and analysis was performed separately for electric and gas projects, where applicable. To provide a comparison with 2013 to 2015, 2010-12 results were weighted by fuel-based MMBtu and all four sets of results are presented in Table 5-7 and Figure 5-1. These exhibits show that 2015 NTGR results for each PA fall within the range of scores seen from the 2010-12 through 2014 evaluations.

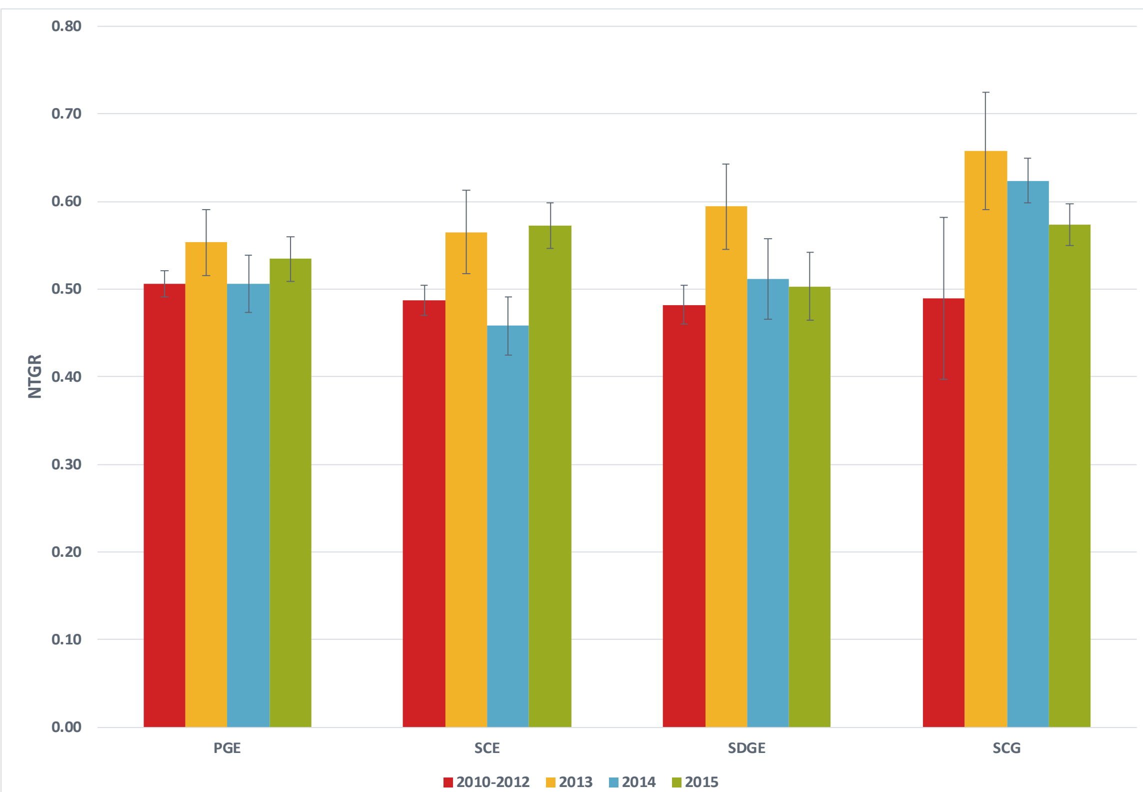
- 2015 NTGRs by PA range from 0.50 to 0.57.
- **PG&E:** 2015 weighted NTGRs results for PG&E (0.53) increased by 6 percent compared to 2014, but no statistically significant differences are observed across the four evaluation cycles.
- **SCE:** 2015 NTGR results for SCE (0.57) are the same as the 2013 results (0.57) but are higher than 2010-12 results (0.49) and 2014 results (0.46).
- **SDG&E:** 2015 NTGR results for SDG&E (0.50) are lower than 2013 results (0.59) but are similar to 2010-12 and 2014 results.
- **SCG:** For SCG the 2015 weighted NTGR across all projects is 0.57. SCG NTG values have declined over the last three evaluation cycles but are still higher than the 2010-12 value of 0.49.

TABLE 5-7: COMPARISON OF 2010-12, 2013, 2014, AND 2015 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	2015 Mean NTGR	2015 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	0.53	0.49 to 0.58
SCE								
MMBtu*	0.49	0.47 to 0.50	0.57	0.52 to 0.61	0.46	0.42 to 0.49	0.57	0.53 to 0.62
SDG&E								
MMBtu*	0.48	0.46 to 0.50	0.59	0.55 to 0.64	0.51	0.47 to 0.56	0.50	0.44 to 0.57
SCG								
MMBtu*	0.49	0.40 to 0.58	0.66	0.59 to 0.73	0.62	0.60 to 0.65	0.57	0.54 to 0.61

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

FIGURE 5-1: COMPARISON OF 2010-12, 2013, 2014, AND 2015 WEIGHTED MMBTU* NTGR RESULTS



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

5.3 NTG SENSITIVITY ANALYSIS

NTG ratios were first calculated for each of the sample points based on equal weighting for each of the three input scores. A sensitivity analysis of the resulting NTGRs was then 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 5-8. 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 5-8: 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)	0.51
2. 50% weight to score 1, 25% to scores 2 and 3	0.51
3. Remove score 1, 50% weight to scores 2 and 3	0.50
4. 50% weight to score 2, 25% to scores 1 and 3	0.53
5. 50% weight to score 3, 25% to scores 1 and 2	0.50
6. Use only score 3 if no-program likelihood is 10	0.52

* Based on simple averaging.

These results indicate that the resulting NTGR results are not very sensitive to the weighting scheme used. 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 are key parts of the NTG survey approach applied in this evaluation.

5.4 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. The analysis was performed for the lowest and highest NTG ratio quartiles,[73] i.e., the group with the lowest NTGRs, corresponding to a threshold value of 0.36 and lower in 2015, and the group with the highest NTGRs, corresponding to a threshold value of 0.70 and higher in 2015. The

[73] Each quartile consists of 25% of project results, and the groupings are assigned based on the NTGR values for the associated projects. Each quartile has identical numbers of projects within a given year.

goal of this analysis was to highlight the factors and characteristics of the groups of projects with both the strongest and weakest levels of program influence. This information not only leads to improved understanding of the results and underlying factors, but also highlights the characteristics of the strongest and weakest groups of projects with respect to program influence. In turn, these characteristics support the development of a set of actionable recommendations regarding strategies to improve program influence going forward.

Table 5-9 presents the results of this analysis for these highest and lowest quartiles, for program years 2013, 2014 and 2015, and across all four PAs. Results are reported for only the percentage of respondents offering the strongest responses (importance scores of 8, 9 or 10). The sample size is noted in the parentheses following the percentage.

The NTGR information in the top four rows of the table characterizes these highest and lowest quartiles, for each year, with respect to quartile definitions (i.e., threshold and mean NTGR values) and numbers of responses for each year. Regarding key NTGR drivers (in the lower section of the table), high percentages of projects in the High quartile illustrate those factors contributing to strong program influence. In contrast, high percentages of projects in the Low quartile are of greatest concern with respect to free ridership (weak program influence). The Key NTG Project Drivers therefore provide insight into the factors that drive high and low 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. The number of observations is also shown in parentheses along with the percentage. By examining these values, some associations between important program and non-program factors and high or low NTG values can be observed.

The following are general themes and observations across these analyses:

- **Program Factors**
  - The **program rebate** was considered highly important by virtually all of those with High quartile NTGRs, while for those with Low quartile NTGRs, it was much less important in general. For those in the Low category, between one-third and one-half of respondents scored the rebate as being highly important.
  - Consistent with the above finding on the **program rebate**, nearly all of those in the High category reported that it brought their project economics within their company's acceptable threshold for decision making. In contrast, less than half of those in the Low category reported this effect, signifying the lower importance of the rebate in general on project decision making.
  - Another factor considered very important by a majority of those in the High category was **technical assistance or studies provided by the program**. Those in the Low category found

little value. To illustrate, in 2015, nearly two-thirds of those in the High category rated this factor highly, while only one-third of those in the Low category considered it important.

  - **Previous program experience** was considered important by a majority of those in the High category during 2013 and 2014, while in 2015, only about one-third of respondents found it to be important. In general, previous experience with the program was not a differentiating factor between those in the High and Low categories during 2015.
  - Other program factors, such as having a **recommendation from program staff**, and input from **program marketing materials**, were seen as relatively unimportant by both groups during all three program years.

- **Non-Program Factors**
  - **Industry standard practice**, **corporate policy** favoring energy efficiency, and following **normal maintenance and equipment replacement policies** were generally more important considerations for those in the Low category compared to those in the High category. These specific non-program motivations often drive project decision making, which in turn correlates with low program influence for those projects.
  - Another differentiating factor between the two groups was related to the **timing of the program's involvement** with the customer. The percentage of those that had made their decision before having any discussions with the program was close to zero for High NTG respondents, and nearly 90% (2015) for Low NTG respondents.
  - Non-program factors such as **improved product quality**, **regulatory compliance** and the **age/condition of the previously-installed equipment** were generally unimportant for both groups.

TABLE 5-9: KEY FACTORS AFFECTING NTGRS FOR ALL PAS[74]

	Highest Quartile NTGR			Lowest Quartile NTGR		
Program Year	**2013**	**2014**	**2015**	**2013**	**2014**	**2015**
Number of Responses	39	50	52	39	50	52
NTGR Threshold value	≥0.67	>0.67	>0.70	<0.38	<0.35	<0.36
Mean NTGR	0.81	0.75	0.80	0.23	0.25	0.18

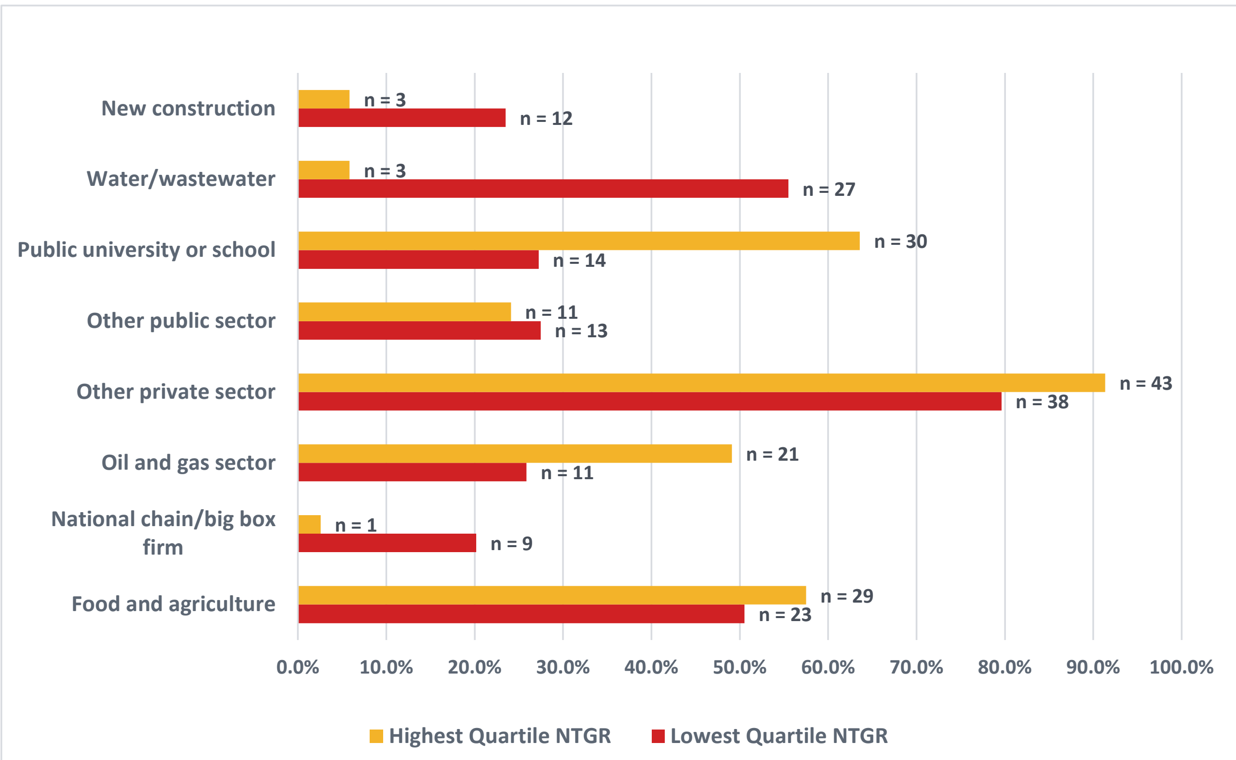
Key NTG Project Drivers	Percent Rating >=8			Percent Rating >=8		
Program Elements						
Program rebate	97% (39)	94% (50)	100% (52)	31% (39)	32% (50)	50% (52)
Rebate brought payback into acceptable range[75]	92% (39)	84% (50)	88% (52)	54% (37)	50% (50)	41% (51)
Recommendation from program staff	21% (39)	35% (49)	16% (51)	21% (39)	26% (50)	25% (52)
Program marketing materials	26% (39)	17% (47)	16% (51)	8% (39)	12% (49)	14% (44)
Program-provided technical assistance or studies	46% (39)	52% (50)	65% (52)	31% (39)	34% (50)	40% (52)
Previous program experience	77% (39)	59% (49)	37% (52)	49% (39)	34% (50)	52% (52)
Non-Program Elements						
Timing						
Made decision before discussions with Program	0% (37)	2.0% (48)	3.9% (51)	54% (37)	70% (46)	88% (48)
Corporate Policy						
Industry standard practice	21% (38)	32% (50)	26% (50)	51% (39)	46% (50)	56% (52)
Corporate policy	41% (34)	57% (49)	33% (52)	54% (39)	62% (50)	48% (52)
Compliance with normal maintenance/replacement	28% (39)	22% (50)	15% (52)	59% (39)	54% (50)	38% (52)
Non-Energy Benefits						
Improved product quality	3% (39)	6% (49)	0% (51)	8% (39)	0% (50)	0% (45)
Environmental Compliance						
Regulatory compliance	0% (39)	0% (50)	0% (52)	5% (39)	0% (50)	2% (52)
Project Context						
Importance of age/condition of old equipment	8% (39)	10% (49)	35% (51)	36% (39)	36% (50)	40% (45)

[74] 2013 data did not include Nonresidential New Construction (NRNC) whole-building projects. For 2014-15, NRNC respondents were not asked about influence of marketing materials, product quality or age or condition of existing equipment.

[75] Self-reported finding.

NTGR trends were also examined by **Market Segment**. Program influence was highest among the Public University/School and Other Private Sector categories. The latter category consisted of a wide range of business types. Influence was lowest for the National chain/Big Box store, Water/Wastewater and New Construction categories. Figure 5-2 below illustrates these relationships based on a graphical comparison of the cumulative market segment shares from 2013 to 2015 for these segments, stratified by the Low and High categories.

FIGURE 5-2: PROGRAM INFLUENCE BY MARKET SEGMENT FOR LOW AND HIGH NTGR CATEGORIES (2013-15)



As noted in Chapter 7, 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.

6 PROJECT PRACTICES ASSESSMENTS

6.1 INTRODUCTION

Project Practices Assessments (PPAs) are structured site-specific reviews of Program Administrator (PA)[76] 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.[77] The PPA results are a companion to the Chapter 4 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.

The PPA process provides impact-oriented findings and feedback to the PAs. The PPA process is conducted on all sampled gross impact points. Previous evaluation cycles bifurcated results based on applications with a customer agreement date falling pre-2013 and 2013+, in attempts 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. For this evaluation cycle, the number of pre-2013 applications were greatly reduced, so no differentiation was made between pre-2013 and 2013+ projects.

---

[76] 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.

[77] 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.

Instead, this third-year evaluation has provided the opportunity to provide meaningful trends over the last three years, in areas of mean rating scores for baseline determination, their calculation methods, and their calculation inputs and assumptions. The evaluation team is also able to show how the percent of ineligible measures or the overturned project types has changed over time, in response to evaluation guidance.

6.2 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.

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 rating scale of 1 to 5 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.

6.3 PROJECT PRACTICES ASSESSMENT RESULTS

This report presents unweighted PPA results by PA. In total, 186 individual records from 148 projects were evaluated across all PAs.[78] This section presents and discusses PA and time period-specific results for the most critical aspects of the PPA, especially those identified in Chapter 4 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.

6.3.1 Project Eligibility Assessments

Table 6-1 presents the PPA findings regarding project eligibility by PA. 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 PA, Table 6-1 displays the number of ineligible and partially ineligible measures. The table also presents the ex-post conclusions for 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 (25 of 70 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, stipulations in the EE policy manual, or other, project-specific reasons. These same issues were identified in Chapter 4 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,

[78] Note that there were actually 189 records associated with the 148 gross M&V points. This number includes two SDG&E records which were removed from the PPA analysis because they are "incentive only" measures that are not subject to engineering review. There was also one additional PG&E record removed which was noted as an "Application Assist" record. There were additional records claimed in the CIS data which were merged together to form a single measure record. Some examples where this happened were when multiple records reflected whole building projects which should not be reported as individual records, or where one record provided the Therms savings and the other record provided the kWh savings for the same measure.

manuals, training, and quality control procedures. Finally, normal maintenance practices at a facility should be reviewed as these are considered ineligible for incentives.

PG&E and SCE were found to have the highest percent of measures which were found to be either ineligible or partially ineligible (and also had the highest occurrence of code/ISP-related issues). Additionally, on average, PG&E had 2.3 reasons why a measure was determined ineligible or partially ineligible, while SCE typically had 2.0. SCG and SDG&E had 1.5 and 1.8 reasons respectively.

TABLE 6-1: SUMMARY OF INELIGIBLE MEASURES, AND EX-POST M&V CONCLUSIONS WHY MEASURES ARE INELIGIBLE

Parameter Examined	**PA Eligibility Treatment***				
	PG&E	**SCE**	**SCG**	**SDG&E**	**Total**
Number of Measures Evaluated	50	57	33	43	183
Number of INELIGIBLE Measures	6	7	5	1	19
Number of Partially INELIGIBLE Measures	7	4	1	4	16
Percent of Measures Found to be INELIGIBLE	12%	12%	15%	2%	10%
Percent of Measures Found to be Partially INELIGIBLE	14%	7%	3%	9%	9%
Evaluation Conclusions Why Measures are INELIGIBLE or Partially INELIGIBLE					
Program rules	2	2	1	0	5
Normal maintenance	1	3	0	0	4
Operating practice change	1	0	0	0	1
CPUC decisions	2	0	0	1	3
CPUC guidance	3	6	1	1	11
Requirement that measures exceed code / ISP baseline	9	10	4	2	25
Previous EAR guidance	0	0	0	1	1
Previous evaluation findings	1	0	1	0	2
Project boundary condition	0	0	0	1	1
EE Policy Manual	1	0	0	0	1
Multiple PA fuels (includes cogeneration and fuel switching)	1	0	0	1	2
Three prong test	0	0	1	0	1
Non-PA fuels and ancillary impacts (i.e., cogen, refinery gas, WHR, etc.)	2	0	0	1	3
Other	7	1	1	1	10

* There are 13 measures included in Table 6-1 that were determined to be ineligible or partially ineligible, but which were identified in the Chapter 4 Discrepancy Analysis results as having alternative explanations for differences between ex-post and ex-ante savings estimates, such as inappropriate baseline. In ten of these cases eligibility issues are associated with measures that fail to exceed baseline energy efficiency levels, as well as other factors related to program rules, operational changes, CPUC guidance and lack of conformance with the Energy Efficiency Policy Manual.

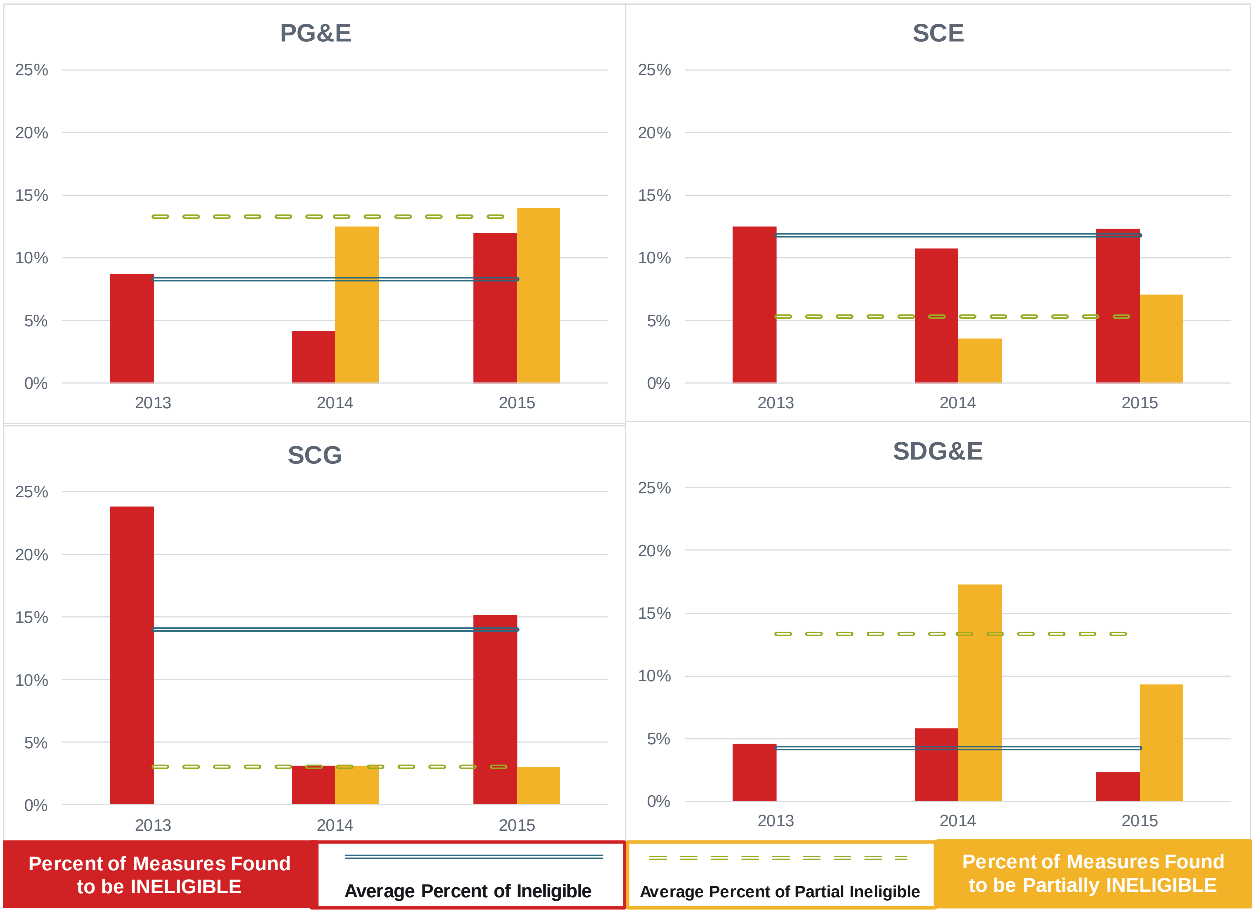
Comparison to Previous Years

The evaluation team also made a comparison to previous years. Figure 6-1 displays how the percent of ineligible or partially ineligible measures have changed over the past three program years.[79] Across three of the four PAs, the evaluation team found that the percent of measures that were determined to be ineligible or partially ineligible, actually increased in 2015, and surpassed the three-year average. Across all PAs, the main reason why measures were considered "ineligible" or "partially ineligible" were due to the requirement that measures exceed either ISP or baseline. This appears to be an ongoing issue, as the 2014 PPA analysis found that this was again the main reason for ineligibility.

One potential reason for the increase in the number of ineligible measures could be related to an inability to keep up with ongoing CPUC decisions and guidance, and previous evaluation findings or EAR guidance. The second-highest reason for ineligibility was due to prior CPUC guidance. As evaluation input has been on-going, prior guidance or findings in different scenarios and unique projects has also increased. PAs may be finding it difficult to keep track of these prior findings and guidance for specific situations. *PAs are responsible for maintaining a central method of tracking IALC findings, EAR findings, and CPUC decisions. PA attention to cataloging and disseminating this information is crucial to PA efforts to improve with respect to underperforming projects, including those with eligibility issues. PAs should also strive to develop checklists to aid in this effort.*

[79] Partial ineligibility was not reported for the 2013 program year.

FIGURE 6-1: COMPARISON OF PROJECT ELIGIBILITY ACROSS PROGRAM YEARS



PA-Specific Findings and Recommendations

The following bullets describe PA-specific findings based on the project eligibility assessment.

- SDG&E had the fewest eligibility issues for the 2015 projects overall, affecting just 11% of the evaluated measures. In addition, it was the only PA to show a decline in the number of measures which were found to be ineligible or partially ineligible over the past three years. In both the ineligible category and the partially ineligible category, SDG&E saw a lower percentage of these measures compared to both the 2014 findings and the overall three-year average findings for SDG&E. When looking at the evaluation conclusions, only two measures cited *Requirement that measures exceed code/ISP baseline*, which was the largest evaluation conclusion category across all PAs. While SDG&E had the best ratings in this category overall, the percent of partially ineligible measures is still on the high end compared with other PAs. This may be partially due to the fact that SDG&E had a higher number of Whole Building projects, where multiple measures were evaluated together rather than on an equipment-basis.

- Although SCG saw the lowest percent of both ineligible and partially ineligible measures in 2014, the number of ineligible measures increased significantly in 2015, with the highest percent increase (12%) for all four PAs. Over all three years, SCG has the highest average percent of ineligible measures at 14%. This may be due, in part, to the fact that SCG had the lowest number of measures evaluated, with only 33 out of the total 183 projects.
- PG&E had one of the highest average percent of partially ineligible measures over the PY2014 and PY2015 evaluation cycles. While the overall average for percent of ineligible measures was on the lower end in 2014, that percentage grew eight points from last year. There were a large number of measures where the evaluation conclusion was noted as the *Requirement that measures exceed code/ISP baseline*, and a high number of conclusions classified as "*other*".[80] Overall, PG&E had the highest percent of measures that were classified as either ineligible or partially ineligible (26 percent).
- SCE did not show much change over the last three years in their percent of ineligible measures, however, the percent of partially ineligible measures did grow slightly from last year. SCE saw the second highest total percent of ineligible or partially ineligible measures, as well as the second highest number of average reasons for ineligibility or partial ineligibility per measure.

In Chapter 4 a total of 30 projects were identified as having zero or negative ex-post MMBtu gross savings, 9 for PG&E, 8 for SCE, 7 for SCG and 6 for SDG&E. Eligibility issues were identified for eleven of those projects, with PG&E, SCG and SDG&E accounting for one or two such instances each, for a total of 5 projects, and SCE accounting for another six projects. For SCE there is considerable room for improvement in eligibility treatment. As was noted in chapter 4, eligibility was the second largest SCE discrepancy factor leading to downward adjustments to ex-ante savings estimates.

6.3.2 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 4 that inappropriate baseline selection was one of the leading causes for downward

[80] There were three reasons which cited Title 24 as the reason for project ineligibility, which would fall under the same category as "Requirement that measures exceed code/ISP baseline", and two measures that cited "previous ISP research conducted by the PA". One measure cited that backup equipment was not eligible for incentives unless operated on a regular basis, and the final heating measure cited that the measure was ineligible because the PA did not provide the fuel to heat the facility.

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 6-2 presents the frequency of ex-ante and ex-post agreement (and disagreement) on project type by PA. These are compared to PY2014 and PY2013 to show how the numbers have changed over time. Across all PAs, the average percent of project types that matched was only 58%. For the PA-specific findings, this ranged by +/- six percent. What is more noticeable however, is the downward trend of the percent of project types which are found to match. Rather than increasing, as would be expected over several years of evaluation, the PA claimed project types are being overturned more and more often.

There are several reasons why this might be the case:

- This Project Type Assessment compares the evaluated project type to the ex-ante project type claimed in the tracking data. However, evaluators also identified ex-ante project type from project documentation. In many cases, the project type inferred or recorded from the project documentation did not match the project type recorded in the tracking data. Thirty-one out of the 183 measures (17%) had a project type which differed in the tracking data and the project documentation. It is possible that in some cases, the ex-ante project type was correctly specified in the project documentation, but that project type didn't make its way into the final tracking data.
- To the extent that project type is also miss-specified in the project documentation, the PAs should strive to carefully examine and document the factors considered when establishing project type. For example, ER, add-on, or system optimization projects should provide clear documentation of the age, condition, RUL, and the capability of the existing equipment to meet service requirements through the RUL period.

TABLE 6-2: FREQUENCY OF EX-ANTE AND EX-POST AGREEMENT ON PROJECT TYPE BY PA AND PROGRAM YEAR

PA	Total Project Types Examined	Project Types Matched	Project Types Overturned	2015 Project Types % Matched	2014 Project Types % Matched	2013 Project Types % Matched
PG&E	50	26	24	52%	73%	80%
SCE	57	36	21	63%	46%	69%
SCG	33	18	15	55%	66%	69%
SDG&E	43	27	16	63%	67%	75%
All PAs	183	107	76	58%	62%	74%

Table 6-3 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. The highest quantity of overturned project types was found to be add-on, replace-on-burnout, and new construction measures. It is notable, however, that add-on and new construction also represent the project type most frequently assigned by the PAs – roughly half of all projects are assigned add-on and roughly one-third of all projects are assigned new construction; the evaluation agrees with the PAs assignment in nearly two-thirds of those projects. ER and ROB measures were overturned in 72 percent, and 71 percent of cases, respectively. System-optimization and Natural Replacement were overturned in every case, but only 12 total measures were specified under these two project baseline types. Ten out of the 15 overturned New Construction projects were changed to Major Renovation or Capacity Expansion projects, which are treated similarly to New Construction projects and still utilize code or ISP technical equipment baselines.

TABLE 6-3: PA VS. EVALUATION SPECIFIED PROJECT TYPE – ALL PAS

			PA-Specified Project Type								
			Add-on	Capacity Expansion	Early Replacement	Major Renovation	New Construction	Natural Replacement	Replace on Burnout	System Optimization	Multiple
Number of measures evaluated (n)			183								
Frequency of PA-Specified Measure Type (n)			79	0	18	0	49	2	21	10	4
Evaluation-Specified Project Type	Frequency of Measure-Level Obs.	(n)									
	Add-on	77	61	0	3	0	3	0	3	6	1
	Capacity Expansion	7	0	0	1	0	3	2	0	0	1
	Early Replacement	14	3	0	5	0	1	0	5	0	0
	Major Renovation	12	2	0	0	0	7	0	2	1	0
	New Construction	35	1	0	0	0	34	0	0	0	0
	Natural Replacement	15	3	0	6	0	0	0	4	2	0
	Replace on Burnout	8	1	0	1	0	0	0	6	0	0
	System Optimization	6	2	0	2	0	0	0	1	0	1
	Multiple	9	6	0	0	0	1	0	0	1	1

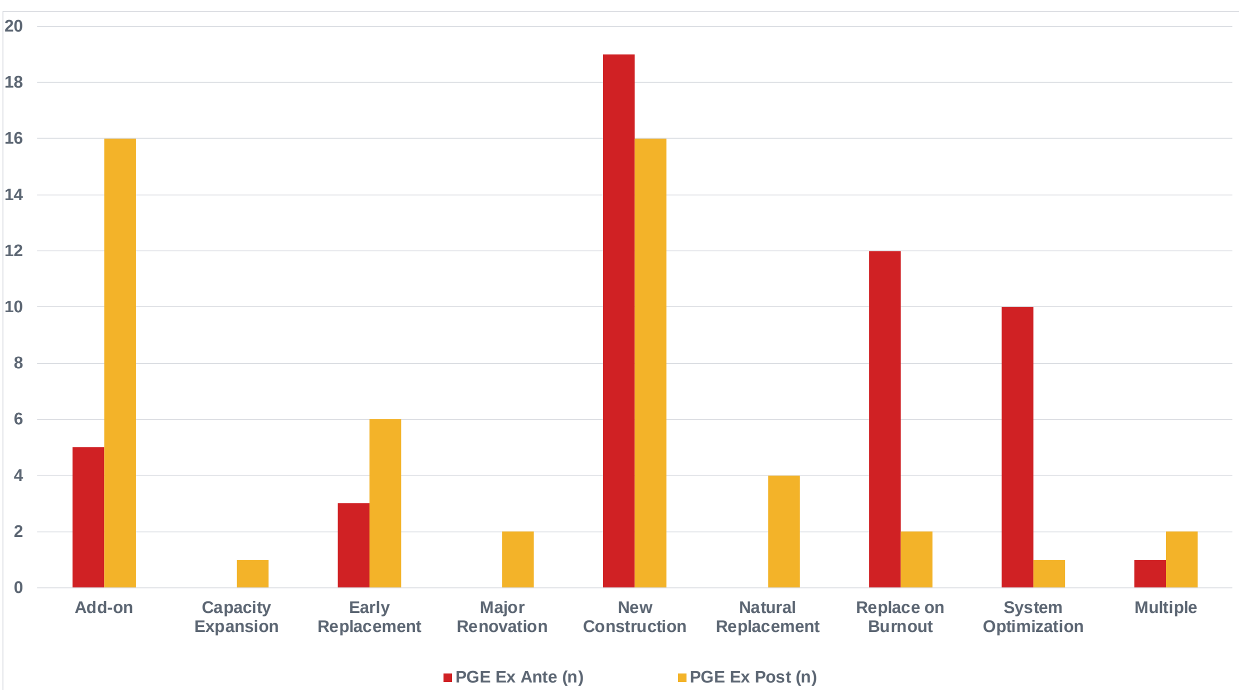
PA-Specific Findings

PG&E

The count of ex-ante and ex-post project types for PG&E are shown below in Figure 6-2. A New Construction project baseline type was most often used by PG&E engineers, followed by Replace-on-burnout and System Optimization. The evaluation team found that more often than not, the Replace-on-Burnout and System Optimization project types were incorrect, and overturned these decisions. In all, there were 24 out of the original 50 ex-ante project types which were overturned. Replace-on-burnout and System Optimization were the most commonly overturned baseline types. Overturned projects were most commonly assigned as Add-On measures (11), Natural Replacement (four), and Early Replacement (four) by the evaluators. A PG&E-specific table, similar to Table 6-3 can be found in Appendix F.

As shown earlier in Table 6-2, the percent of project types which match between the ex-ante and evaluation findings have decreased over the years. PG&E had the highest number of matching project types in 2013 (80%), but now have the lowest number out of all four PAs (52%). One possible reason for this may be the high number of mismatched project types between what was inferred from project documentation and what was recorded in the tracking data (15). If PG&E is more diligent about ensuring that the correct project type is recorded in both the project documentation and the tracking data, it is possible that they will see an increase in the number of project types that match in future program years.

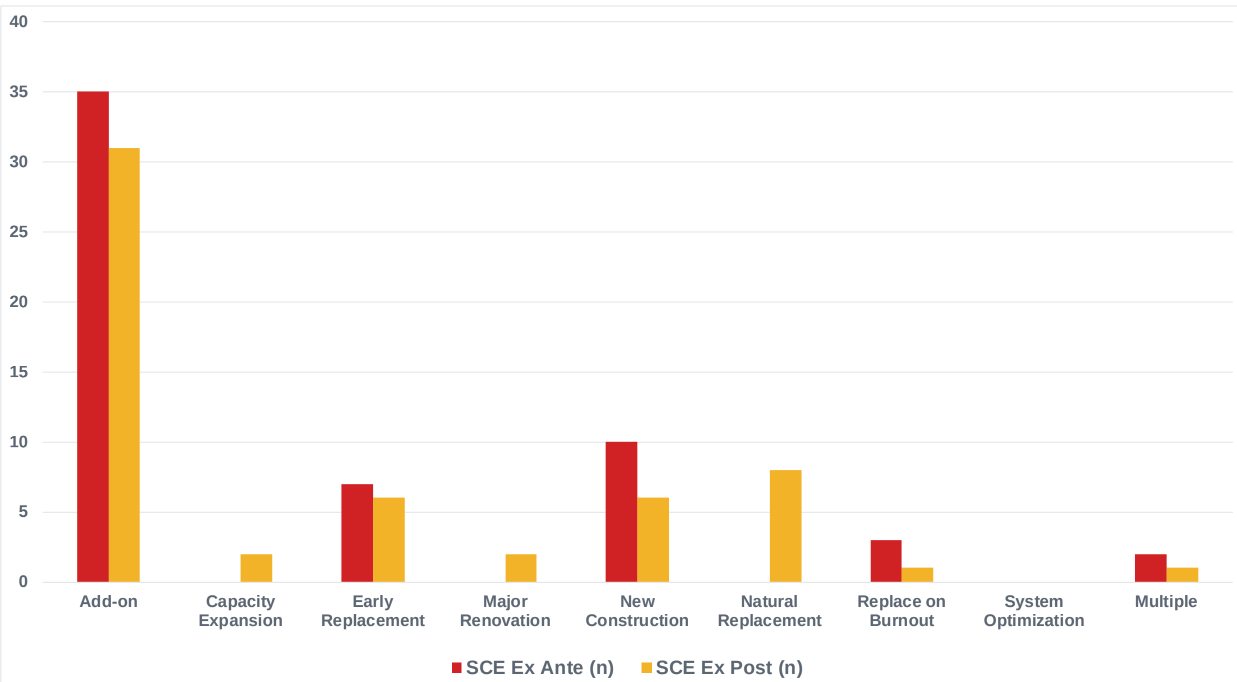
FIGURE 6-2: PG&E COUNT OF PROJECT TYPES – EX-ANTE AND EX-POST



SCE

The count of ex-ante and ex-post project types for SCE are shown below in Figure 6-3. Overall, the percent of project types which matched has increased 17 percent from 2014 and was four percent higher from its three-year average, as shown above in Table 6-2. SCE was tied for the highest percentage of project types which matched out of the four PAs, at 63 percent. SCE was found to have the highest number of Add-on measures (31), which was followed far behind by eight Natural Replacement measures. Natural Replacement was not selected as a project type by ex-ante engineers for any sampled SCE project. Seventy-one percent of the Early Replacement ex-ante claims were overturned, although this is obscured in the chart by four other claims which the evaluation team changed to Early Replacement. Similarly, all of the ex-ante Replace-on-Burnout claims were overturned. These SCE-specific findings can be found in Appendix F, in a table similar to Table 6-3.

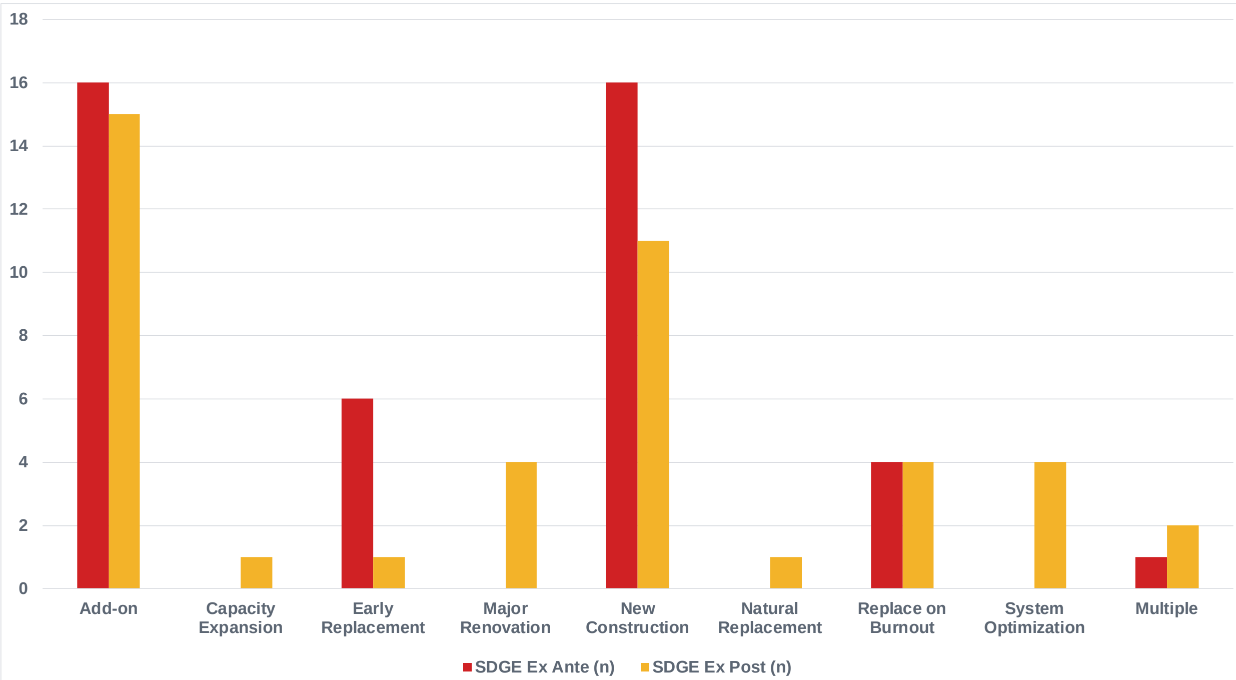
FIGURE 6-3: SCE COUNT OF PROJECT TYPES – EX-ANTE AND EX-POST



SDG&E

SDG&E is tied for the highest number of project types which matched between the ex-ante and ex-post analyses. These results can be seen in Table 6-2. However, the percent of project types which match have still been declining steadily over the past three years. The evaluation team found that SDG&E's PY2015 results were six percent below its three-year average. Like the other PA's, Add-on measures were common for SDG&E, but there were also many New Construction measures. Five of the New Construction measure claims were overturned,[81] and 100% of the ex-ante Early Replacement claims were overturned, although one Add-on claim was changed to Early Replacement by the evaluation team. An SDGE-specific table, similar to Table 6-3 can be found in Appendix F.

FIGURE 6-4: SDG&E COUNT OF PROJECT TYPES – EX-ANTE AND EX-POST



SCG

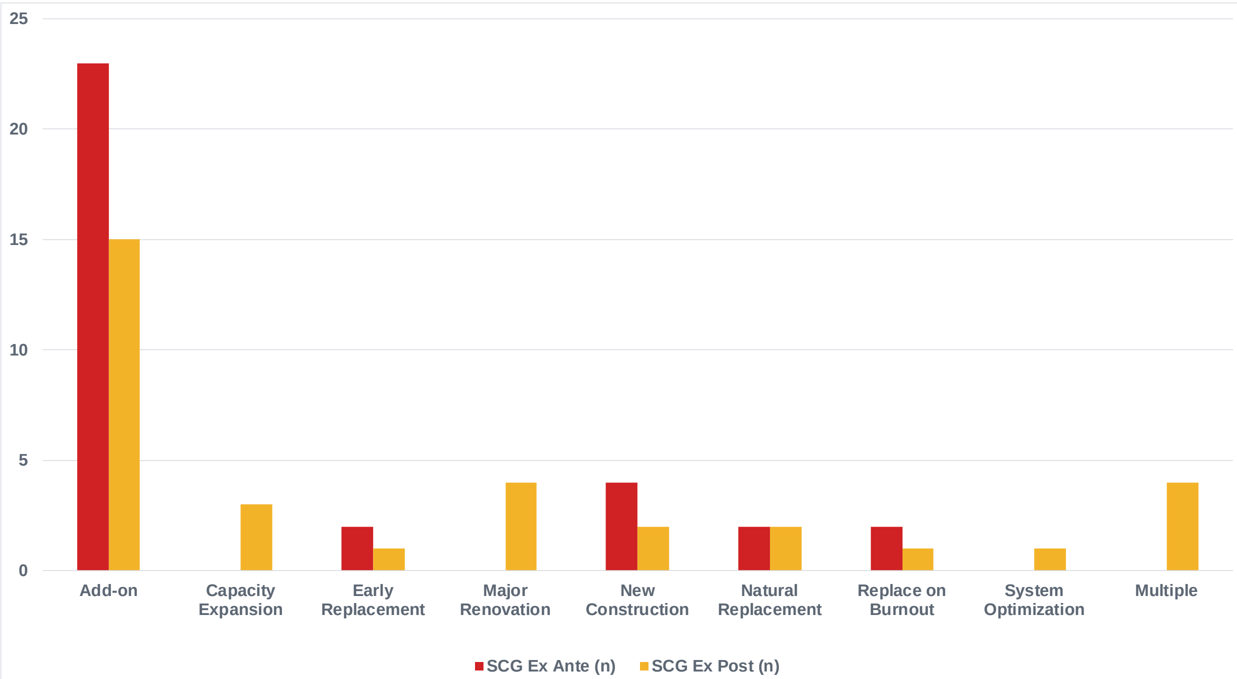
Figure 6-5 below displays the count of project types, both ex-ante and ex-post for the measures in the evaluation sample. As with the other PAs, Add-on measures were identified most often, in both the ex-ante and ex-post. SCG has seen a steady decline in the percent of project types which have matched between the ex-ante and ex-post analyses. SCG's PY2015 findings were 8 percent lower than the three-year average, as seen in Table 6-2. There were eight measures which were initially claimed as Add-on measures but overturned. Of all overturned projects, four were found to be Major Renovation, four more to have Multiple project types, and three more overturned to Capacity Expansion projects. An SCG-specific table, similar to Table 6-3 can be found in Appendix F.

SCG was only found to have three projects out of the evaluation sample where the project type claimed in the tracking data did not match the project type inferred from the project documentation – the lowest

[81] Four of these were overturned to Major Renovation projects, which still use a code or ISP technical equipment baseline.

percent out of all PAs. This number is still at almost 10 percent, but overall SCG has done a good job at ensuring the ex-ante determinations have made their way into the tracking data.

FIGURE 6-5: SCG COUNT OF PROJECT TYPES – EX-ANTE AND EX-POST



6.3.3 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 6-4 presents the frequency of ex-ante and ex-post agreement (and disagreement) on project baseline by PA. Across all PAs in 2015, the percent of project baselines which matched between the ex-ante and ex-post analysis was up from PY2014 but slightly below the three-year average. PG&E and SCG match percentages have declined relative to 2013 and 2014 while SDG&E's 2015 project baseline selection has shown improvement. SCE has seen an increase over PY2014 but a decline relative to PY2013.

In Chapter 4 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 with CPUC baseline policies.

TABLE 6-4: FREQUENCY OF EX-ANTE AND EX-POST AGREEMENT ON PROJECT BASELINE BY PA AND PROGRAM YEAR

PA	Total Project Baselines Examined	Project Baselines Matched	Project Baselines Overturned	2015 Project Baselines % Matched	2014 Project Baselines % Matched	2013 Project Baselines % Matched
PG&E	50	29	21	58%	75%	75%
SCE	57	43	14	75%	63%	80%
SCG	33	22	11	67%	75%	69%
SDG&E	43	38	5	88%	71%	75%
All PAs	183	132	51	72%	70%	75%

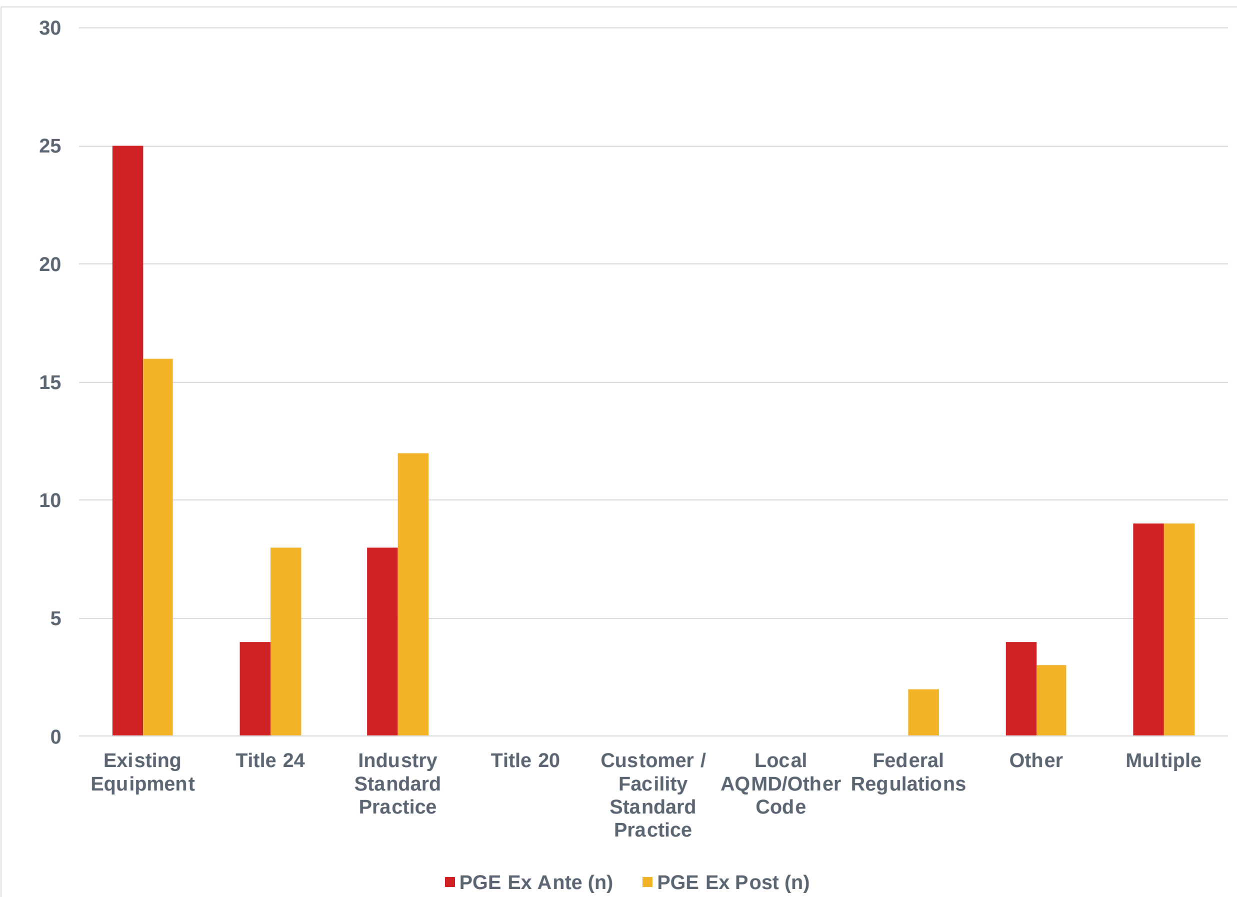
Table 6-5 presents results detailing ex-ante versus ex-post project baseline designations for all PAs. The green shaded cells along the diagonal indicate the number of measures that showed agreement between the PA and ex-post evaluation. Values in the red shaded cells are measures where the project baseline was reassigned by the evaluator. Sixty-two percent of the PA-specified project baselines were specified as existing equipment. This was overturned 24 percent of the time. Results for other baseline types are somewhat mixed across the PAs, yet existing equipment holds the majority of all PA-specified project baselines across all 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 6-5: PA VS. EVALUATION SPECIFIED PROJECT BASELINE – 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
Number of measures evaluated (n)			183								
Frequency of PA Specified Baseline (n)			115	22	17	3	1	0	0	5	20
Evaluation-Specified Project Baseline	Frequency of Measure-Level	(n)									
	Existing equipment	94	87	0	6	0	0	0	0	0	1
	Title 24	31	5	22	0	0	0	0	0	0	4
	Industry standard practice	24	6	0	10	0	1	0	0	3	4
	Title 20	3	0	0	0	3	0	0	0	0	0
	Customer/facility std.	0	0	0	0	0	0	0	0	0	0
	Local AQMD/other code	2	2	0	0	0	0	0	0	0	0
	Federal regulations	2	2	0	0	0	0	0	0	0	0
	Other	10	7	0	0	0	0	0	0	1	2
	Multiple	17	6	0	1	0	0	0	0	1	9

Figure 6-6 through Figure 6-9 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. The overturning of this particular baseline has been a common theme across custom evaluations going back at least a decade or more. Twenty-eight of the original 115 existing equipment claims were overturned. These overturned baseline types were rather evenly distributed between Title 24, industry standard practice, "other", and "multiple." [82] In all 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 local codes and federal regulations. Across all PAs, ex-ante baselines were most commonly overturned to industry standard practice.

FIGURE 6-6: PG&E DISTRIBUTION OF PROJECT BASELINES – EX-ANTE AND EX-POST



[82] 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 6-7: SCE DISTRIBUTION OF PROJECT BASELINES – EX-ANTE AND EX-POST

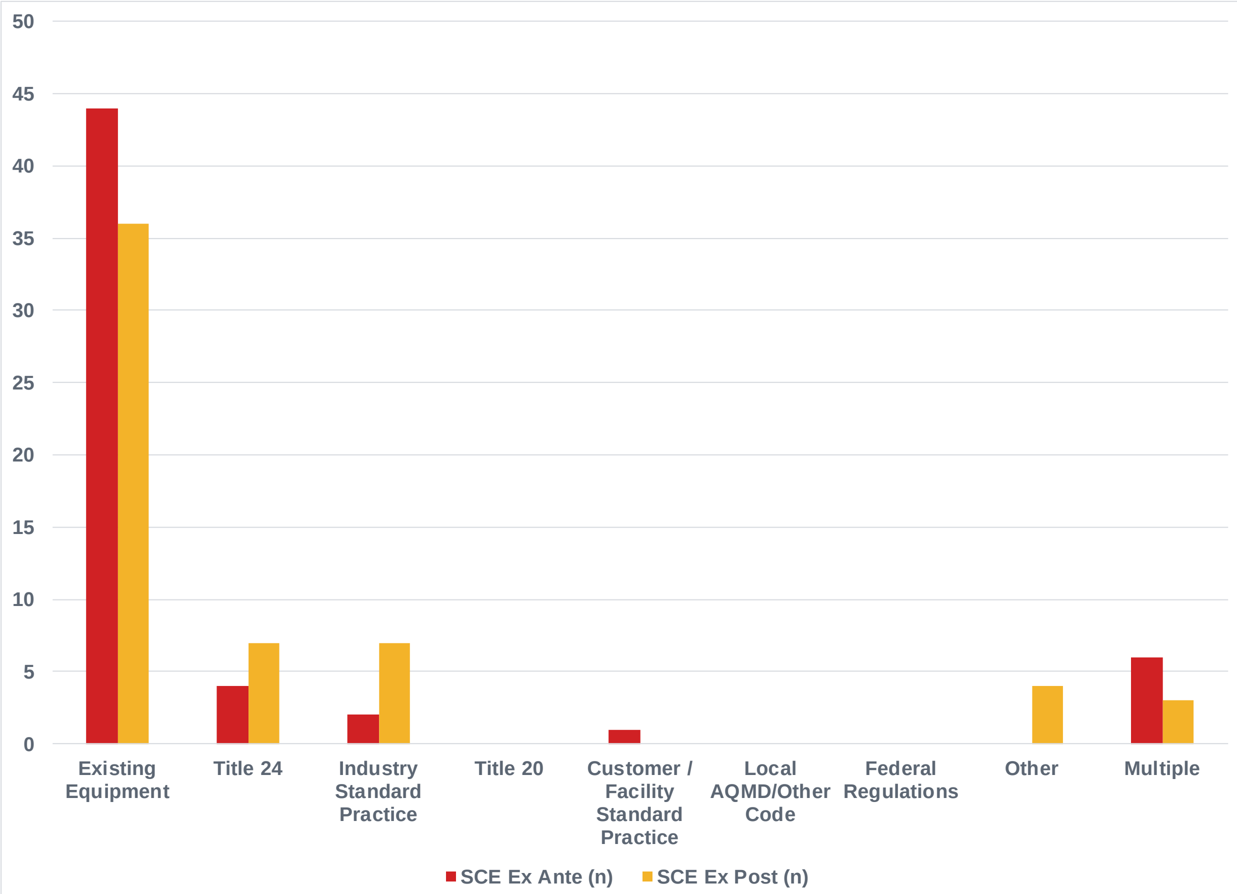


FIGURE 6-8: SDG&E DISTRIBUTION OF PROJECT BASELINES – EX-ANTE AND EX-POST

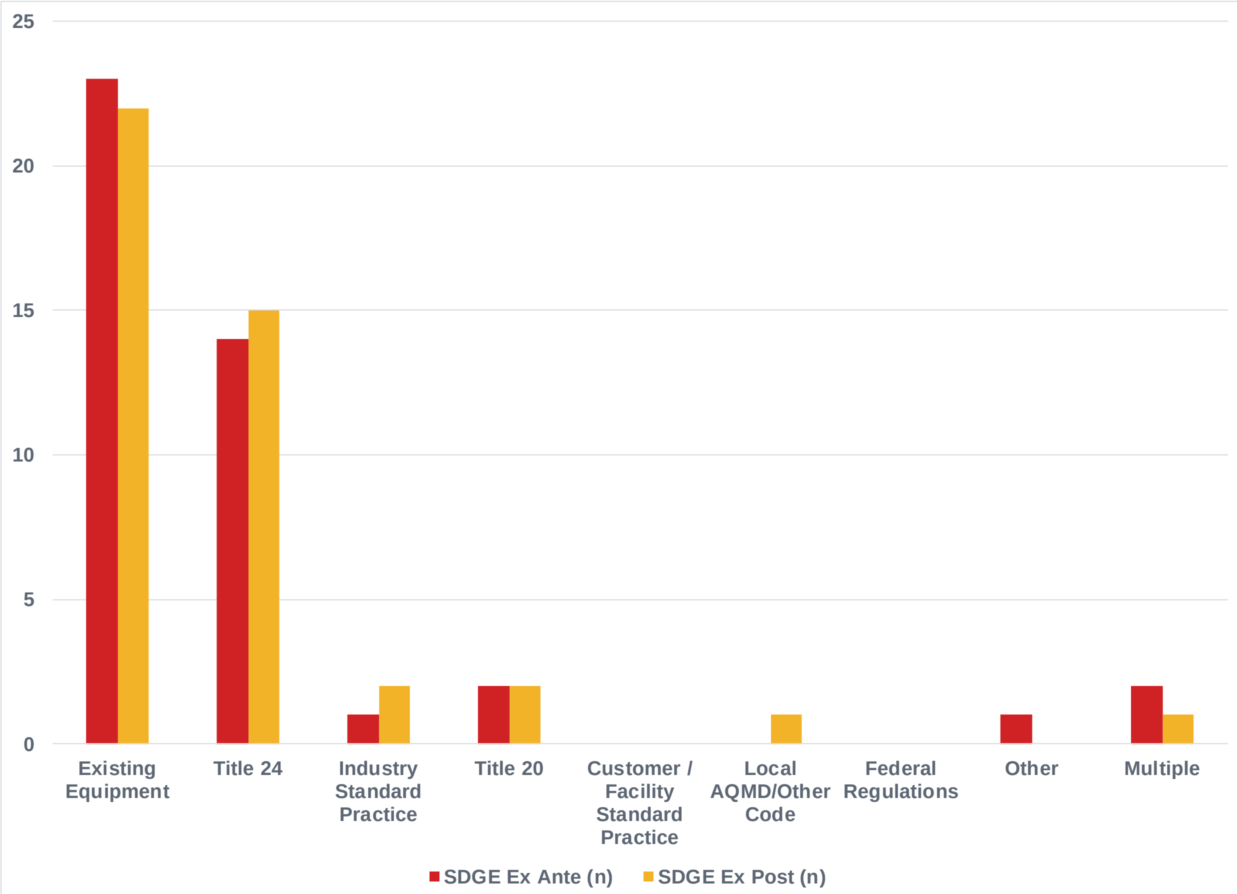
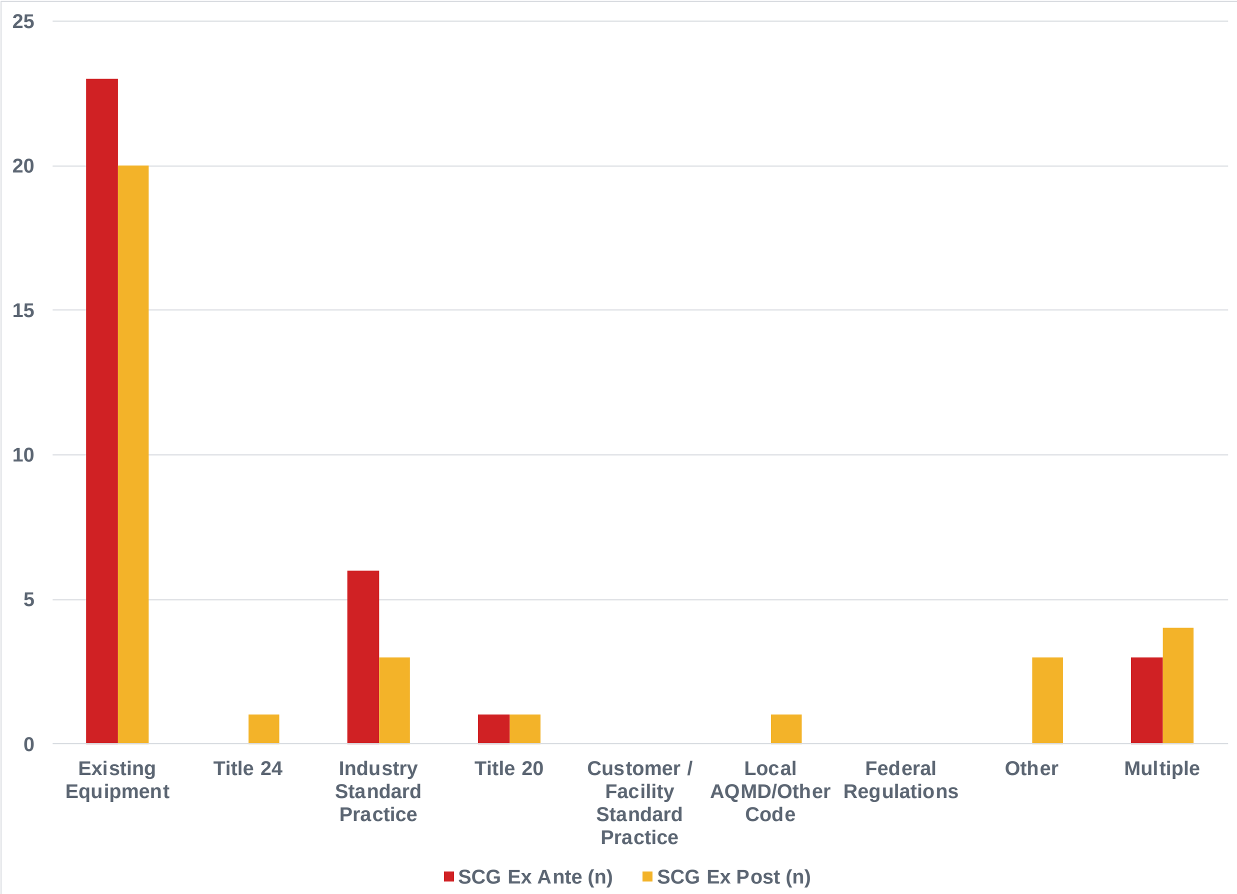


FIGURE 6-9: SCG DISTRIBUTION OF PROJECT BASELINES – EX-ANTE AND EX-POST



6.3.4 Project Baseline Ratings

Table 6-6 and Figure 6-10 (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.[83] For example, ER, add-on, or system optimization projects should provide clear documentation of the age, condition, RUL, and the capability of performance through the RUL of the existing equipment. Other relevant considerations include examination of facility and industry standard practices, applicable codes and standards, and maintenance records.

[83] 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.

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 for the second baseline assignment. Low baseline appropriateness scores generally correspond to the overturned baselines demonstrated above in Table 6-5.

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 6-6: PROJECT BASELINE DOCUMENTATION QUALITY BY PA

Parameter Examined			PA Project Baseline Treatment (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)			
			PG&E	SCE	SCG	SDG&E
Project Baseline	Quality of Documentation Rating	n	50	57	33	43
		Mean	2.28	1.96	2.27	2.23
		Median	2	2	2	2
	Appropriateness of Baseline Determination Rating	n	50	57	33	43
		Mean	2.40	2.30	2.76	2.42
		Median	3	3	3	3
	Quality of Baseline Description Rating	n	50	57	33	43
		Mean	2.98	2.54	2.58	2.26
		Median	3	3	3	2

FIGURE 6-10: MEAN PROJECT BASELINE DOCUMENTATION QUALITY RATINGS BY PA

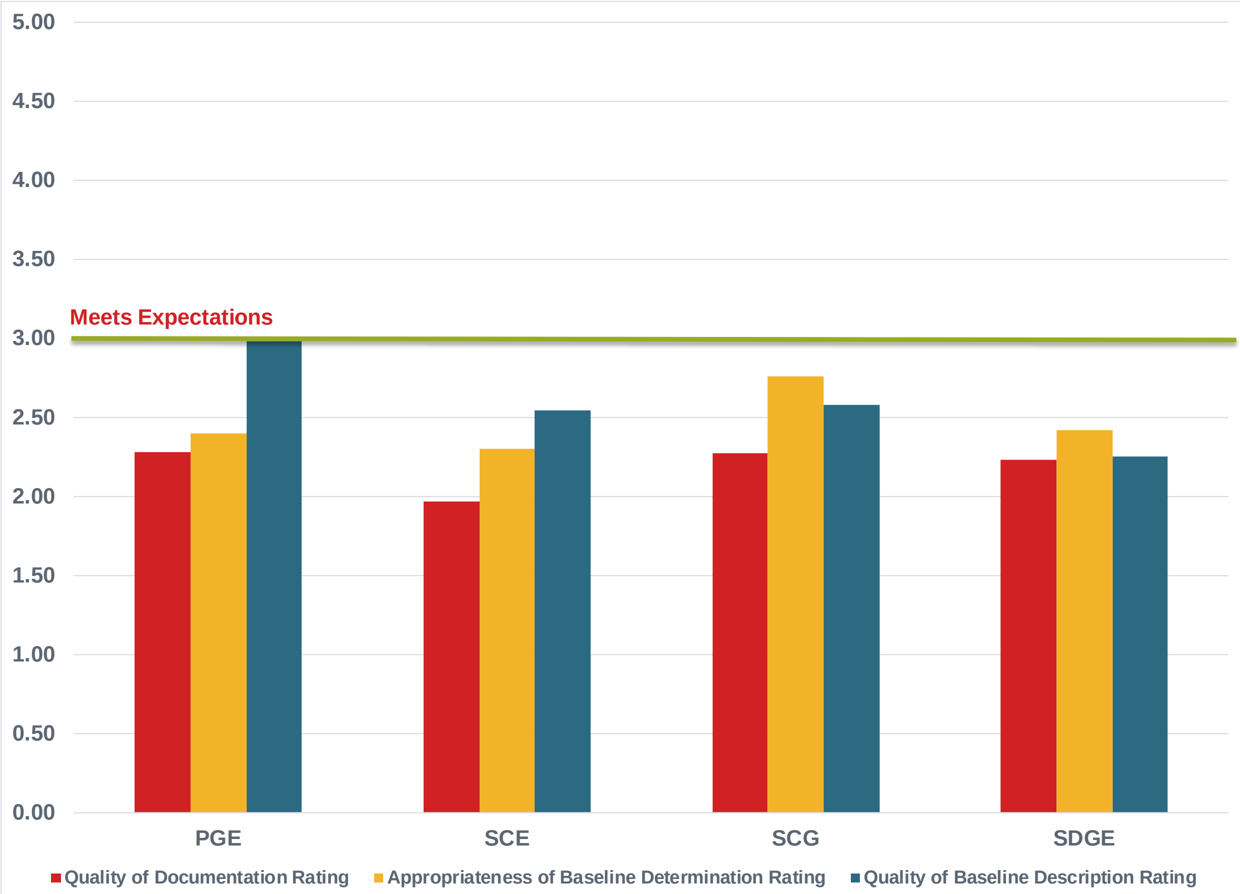


Table 6-6 and Figure 6-10 show that on average, none of the PAs "met expectations" (i.e. mean score of 3). This is true for all three metrics, although PG&E effectively "met expectations" with the quality of baseline description rating. Scores range from 1.96 to 2.98. It should be noted that the majority of scores have medians of 3 (with the exception of the quality of documentation rating), 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.

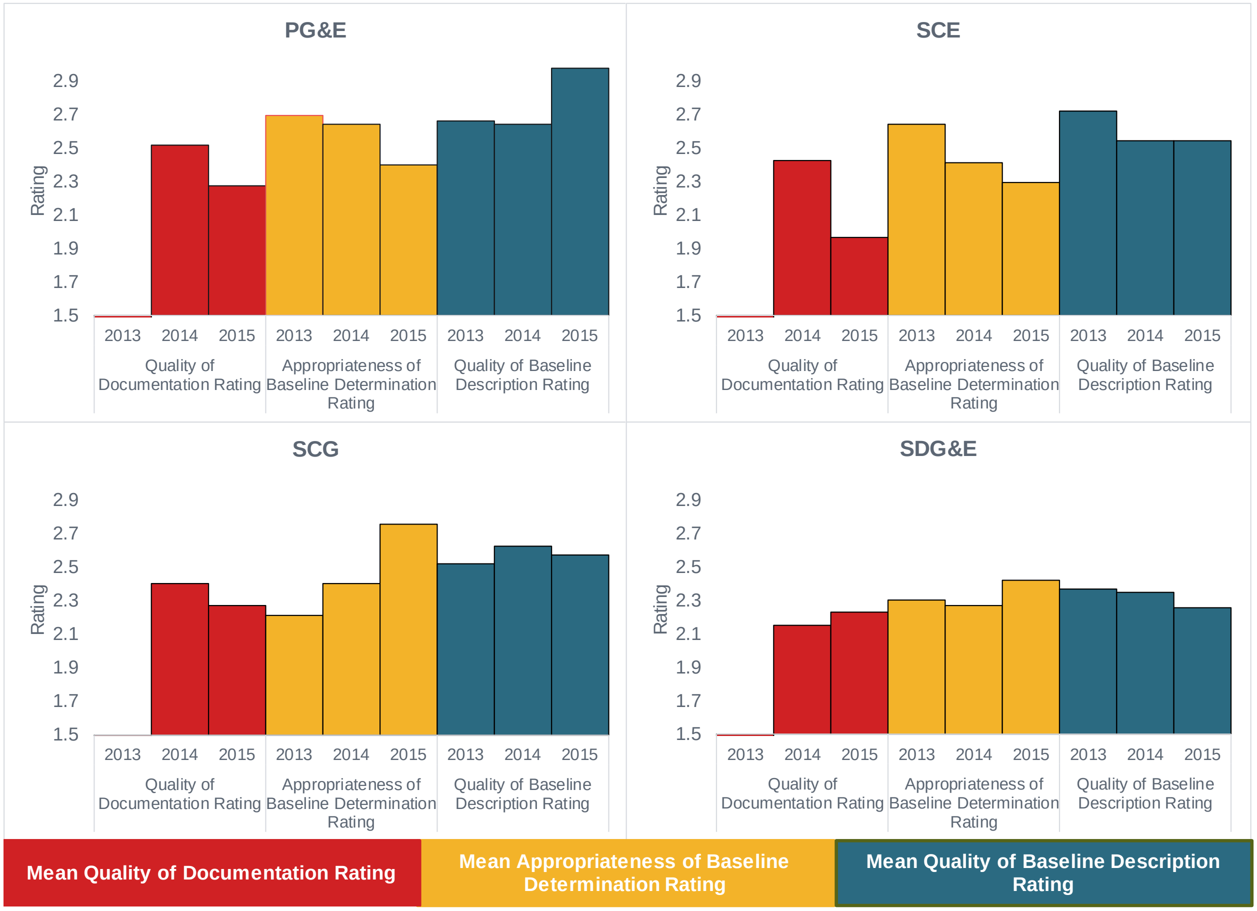
In instances where the documentation quality scores did not meet expectations (scores of 1 or 2), the evaluation most often cited the condition of existing equipment, the age of the existing equipment, the capability of existing/baseline equipment to meet facility service needs, normal facility practices, applicable codes and standards, and the EUL of the equipment when correctly establishing the baseline. For the most part, these were the same items cited in the PY2014 evaluation. 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 applicable codes and standards, which were also common reasons for overturned project baselines.

Figure 6-11 compares the mean project baseline ratings for each PA across the three program years. Across the three metrics for all four PAs, the Quality of Baseline Description was typically found to be the highest, with a mean rating of 2.57 across the three years, while the Quality of Documentation was the lowest, with a mean rating of 2.28 across the two years.[84] Several findings from this annual comparison are highlighted here:

- The Quality of Documentation rating was not only the lowest across the three metrics, but it also was found to decline for three out of the four PAs, with only SDG&E seeing a slight increase in the mean rating from PY2014 to PY2015. Adequate documentation is the only method of supporting ex-ante findings and can be a major factor in the overturning of project types and project baselines, especially when the evaluation team's findings differ from the ex-ante findings.
- The Appropriateness of Baseline Determination rating was found to decrease for both PG&E and SCE, while SCG and SDG&E have seen a rise in their ratings over their three-year mean.
- The Quality of Baseline Description rating was found to increase over the three-year mean for PG&E only, while SCG sat directly at its three-year mean, and SCE and SDG&E were found to be below it.

[84] Quality of Documentation rating was not provided in the PY2013 evaluation.

FIGURE 6-11: COMPARISON MEAN PROJECT BASELINE RATINGS BY PA ACROSS PROGRAM YEARS



* Note that the scale has been adjusted to range between 1.5 and 2.9 so that the trend over the three program years can be more easily identified.

** The Quality of Documentation rating was not analyzed in PY2013.

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. Additionally, Appendix E includes a detailed description of the PPA scoring criteria. The PAs are encouraged to thoroughly examine the attributes contributing to scores of 3, 4, and 5 and include those elements when determining and documenting project types and project baselines.

6.3.5 EUL Assessment

Table 6-7 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 100 percent of measures. By comparison, EUL was not documented as often in the project application files, ranging from 48 percent to 85 percent of measures. While these scores are generally higher than those in the 2014 evaluation (where the range was 13 percent to 83 percent), all PAs are still providing insufficient EUL data in the project files. The mean documentation score ranged between 1.79 and 2.36 for all PAs. A rating of one means an EUL has not been assigned in the project documentation. A rating of two indicates that the EUL is incorrectly claimed, and a rating of three notes that the EUL from the project documentation matches DEER. For measures where a DEER EUL is not available, PAs receive a score of four or five if they provide clear, documented evidence for a reliable EUL. Based on the mean ratings shown below, there were a majority of measures where the EUL was found to be incorrectly claimed.

TABLE 6-7: EUL ASSESSMENT BY PA

Parameter Examined	PA EUL Documentation (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)			
	PG&E	SCE	SCG	SDG&E
Summary of Evaluation EUL Treatment				
Number of Measures Assessed	50	57	33	43
Number of Measures with PA Tracking System EUL Populated (n)	50	57	33	43
Percent of Measures with PA Tracking System EUL Populated	100%	100%	100%	100%
Number of Measures with PA EUL Documented in the Project Application Files	25	45	28	22
Percent of Measures with PA EUL Documented in the Project Application Files	48%	79%	85%	51%
Mean PA EUL Documentation Score	1.84	2.28	2.36	1.79
Median PA EUL Documentation Score	2	2	2	2
Summary of EUL Differences				
Number of Measures with Evaluation EUL Different Than PA Tracking EUL (n)	27	32	14	23
Percent of Measures with Evaluation EUL Different Than PA Tracking EUL	54%	56%	42%	53%
Mean Evaluation EUL (where differences exist)	8.72	7.80	12.28	9.22
Median Evaluation EUL (where differences exist)	6.7	5.86	10	5.23
Mean PA Tracking System EUL (where differences exist)	11.45	11.97	13.43	12.80
Median PA Tracking System EUL (where differences exist)	11	15	14.5	15

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 with only 48 percent of their measures having a documented EUL. PG&E's low mean EUL documentation score of 1.84 largely reflects the missing EUL documentation and, to a lesser extent, EUL's that were documented but incorrect (54 percent).
- SCE saw an increase over PY2014 in the percent of measures with an EUL documented in the Project Application files, with 79 percent of records documented. Over half of the measures (56 percent) were found to have an evaluated EUL that differed from the claimed EUL. These factors led to an EUL documentation score of 2.28.
- SCG's rate of EUL documentation was the highest of the four PAs, at 85 percent, exceeding the PY2014 finding. SCG also had the lowest percent of measures found to have an evaluation EUL that differed from the claimed EUL, at 42 percent. These findings led to the highest mean EUL documentation score out of the four PAs, at 2.36.
- SDG&E's rate of documentation of EUL in the project application files declined from the PY2014 evaluation to 51 percent. The rate of incorrect EUL documentation was also over half, at 53 percent. This led to low mean documentation score of 1.79.

Of the 121 measures that had EUL documented in the project application files, 38 percent were sourced from DEER. No source was provided for 30 percent of the EULs documented in the project applications. The primary source of evaluation-sourced EULs (52 percent) was determined to be "Other", followed by 45 percent sourced from DEER.

Across all PAs the mean evaluation-sourced EULs were lower than the mean tracking system EULs. For the 52 percent of measures that had different PA tracking and evaluation-sourced EULs (a total of 96 out of 183 evaluated measures across all PAs), the mean differences between the evaluation-sourced and tracking EULs for each PA are as follows:

- PG&E: -2.73 years
- SCE: -4.17 years
- SCG: -1.15 years
- SDG&E: -3.57 years
- Across all PAs: -2.91 years

As noted in Chapter 4 LC MMBtu GRR results were lower than FY GRR results (for all PAs, with SCE showing the largest differential) and this EUL difference was *a key factor* driving down the LC GRR results.

6.3.6 Calculations Assessment

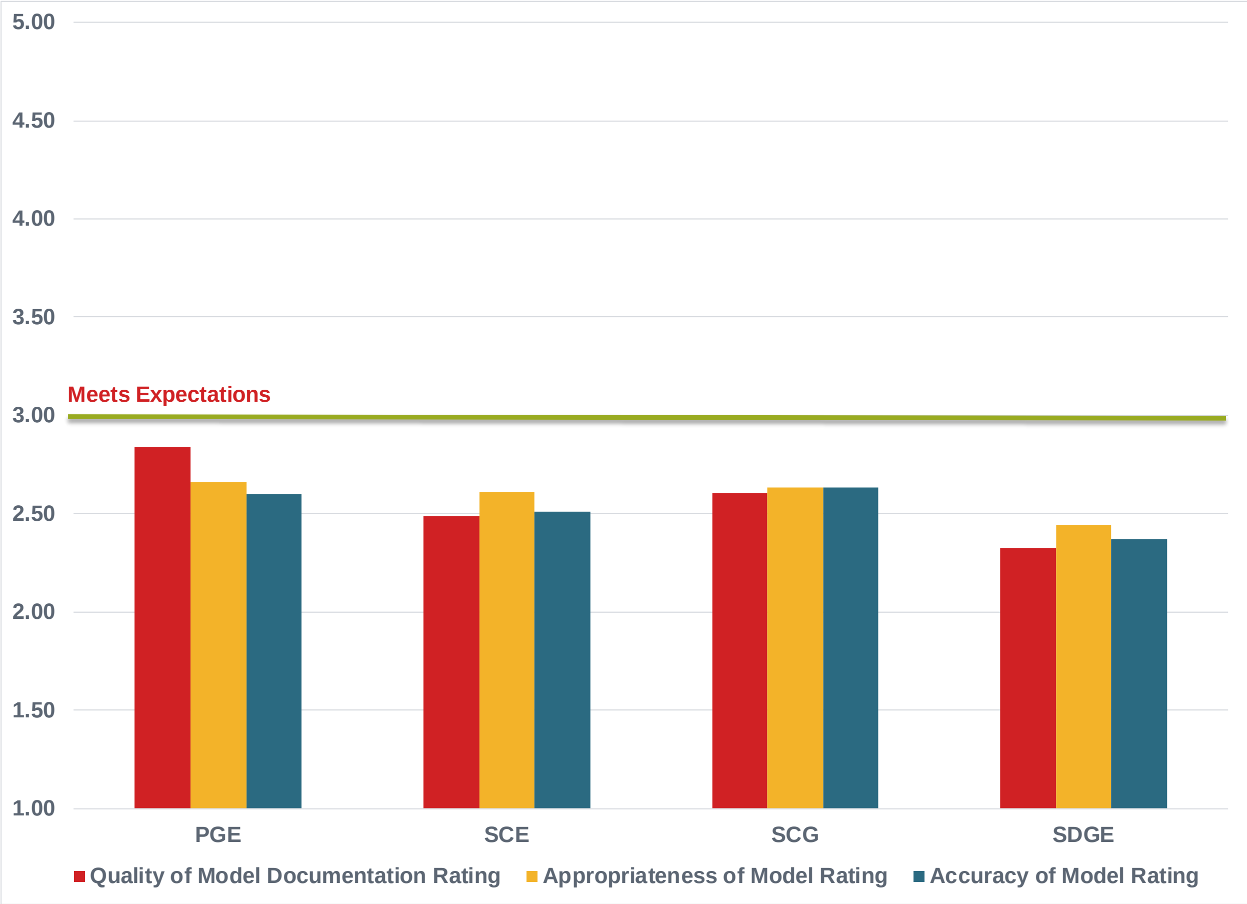
Table 6-8 and Figure 6-12 (graphical representation) below, provide an assessment of the documentation quality, appropriateness, and the accuracy of the PA models in determining measure savings. The Quality of Documentation score reflects the degree to which the PA calculation model is clearly documented 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 6-8: CALCULATIONS METHODS ASSESSMENT BY PA

Parameter Examined			PA Calculation Methods Treatment (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)			
			PG&E	**SCE**	**SCG**	**SDG&E**
Calculation Methods	Quality of Model Documentation	n	50	57	33	43
		Mean	2.84	2.49	2.61	2.33
		Median	3	3	3	2
	Appropriateness of Model	n	50	57	33	43
		Mean	2.66	2.61	2.64	2.44
		Median	3	3	3	2
	Accuracy of Model	n	50	57	33	43
		Mean	2.60	2.51	2.64	2.37
		Median	3	3	3	2
Assessment of Evaluation Use of PA Inputs						
Number of Measures Assessed (n)			45	55	32	40
Evaluation used a different model			38%	33%	28%	28%
Evaluation used a similar model			31%	40%	22%	38%
Evaluation adjusted the PA model			31%	27%	50%	35%

FIGURE 6-12: CALCULATIONS METHODS ASSESSMENT BY PA



The general trend seen in the table and figure above show an approximate average rating across all three metrics and four PAs of 2.5, indicating that the calculation methods for these PAs were generally appropriate, accurate, and well documented. PG&E was on the higher end of the spectrum across all three metrics, while SDG&E was on the lower end of the spectrum across all three metrics.

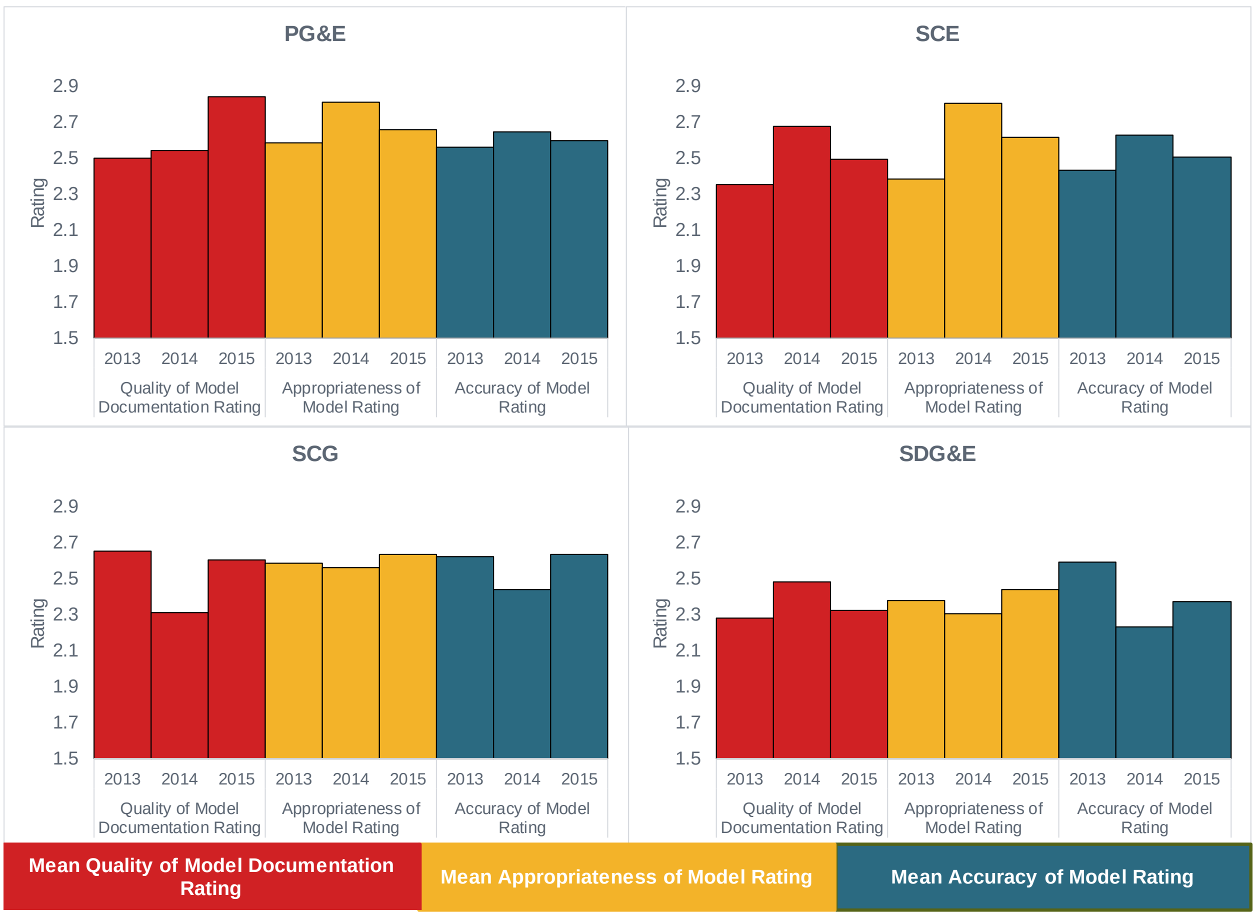
Finally, Table 6-8 shows that the evaluator only used the PA model (or similar model) between 22 percent and 40 percent of measures. In all other cases, the evaluator used an entirely different model or deemed it necessary to make adjustments to the PA models.

Figure 6-13 shows a comparison of the Calculation Method ratings across the three program years.

- PG&E shows a significant increase in their Quality of Model Documentation rating, while the Appropriateness of Model rating dipped slightly down closer to the three-year average, and the Accuracy of Model rating has been rather consistent, +/- 0.1 rating point over the three years.

- SCE's rating score for all three metrics has dropped from PY2014, but increased from PY2013, putting it just around the three-year average.
- SCG has seen an increase in all metrics relative to PY2014, and in all cases except the Quality of Model Documentation rating, are higher than PY2013 as well.
- SDG&E has seen slight increase in their Appropriateness of Model rating over the three years. Their Quality of Model Documentation rating is down from both PY2014 and below the three-year average, while their Accuracy of Model rating is below the three-year average but above PY2014.

FIGURE 6-13: COMPARISON MEAN CALCULATION METHOD RATINGS BY PA ACROSS PROGRAM YEARS



* Note that the scale has been adjusted to range between 1.5 and 2.9 so that the trend over the three program years can be more easily identified.

The overall result from this three-year comparison sees some ups- and downs- in the model ratings, but there is no clear trend one way or the other across any PA. There is no evidence of systematic changes made to address any inadequacies identified in the models. *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*.

6.3.7 Inputs and Assumptions Assessment

Table 6-9 and Figure 6-14 (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 6-9: INPUTS AND ASSUMPTIONS ASSESSMENT BY PA

Parameter Examined			PA Inputs and Assumptions Treatment (1 = Does not meet basic expectations, 5 = Consistently exceeds expectations)			
			PG&E	SCE	SCG	SDG&E
Inputs and Assumptions	Quality of Input and Assumption Documentation	n	50	57	33	43
		Mean	2.60	2.39	2.67	2.35
		Median	3	2	3	2
	Comprehensiveness of Inputs and Assumptions	n	50	57	33	43
		Mean	2.68	2.46	2.58	2.37
		Median	3	3	3	2
	Accuracy of Inputs and Assumptions Rating	n	50	57	33	43
		Mean	2.36	2.35	2.58	2.28
		Median	2	3	2	2
Assessment of Evaluation Use of PA Inputs						
Number of Measures Assessed (n)			46	55	32	42
Evaluation used different inputs			37%	35%	31%	26%
Evaluation used similar inputs			28%	42%	28%	36%
Evaluation adjusted the PA inputs			35%	24%	41%	38%
Assessment of Evaluation Use of PA Assumptions						
Number of Measures Assessed (n)			45	57	29	41
Evaluation used different assumptions			44%	39%	31%	37%
Evaluation used similar assumptions			27%	44%	34%	44%
Evaluation adjusted the PA assumptions			29%	18%	34%	20%

FIGURE 6-14: INPUTS AND ASSUMPTIONS ASSESSMENT BY PA

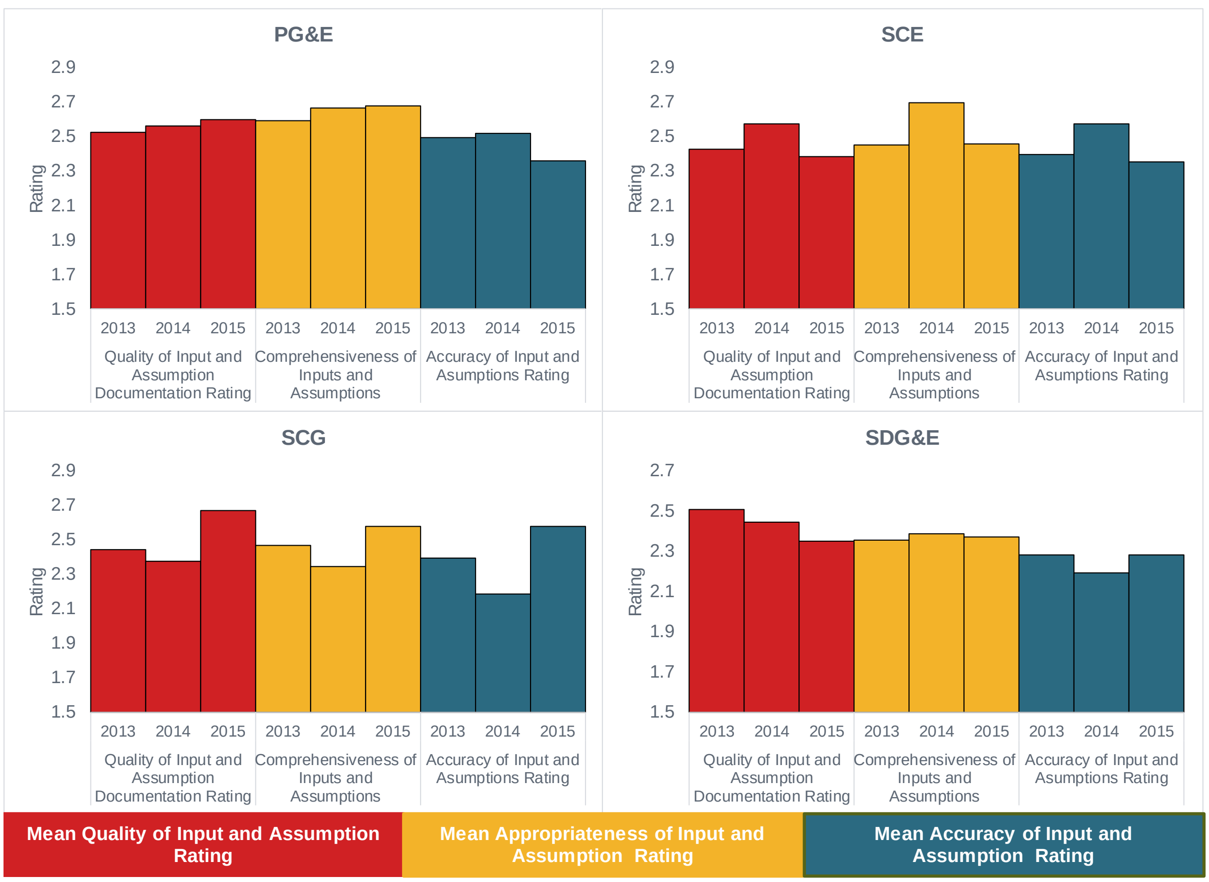
As shown in the table, the mean documentation, comprehensiveness, and accuracy ratings for each PA are between 2.0 and 3.0, indicating that, on average, the PAs fell short of minimum expectations. Additionally, Figure 6-15 shows a comparison of the ratings over the three-program years across these three metrics, by PA. PA specific results are summarized as follows:

- PG&E's average ratings for the three inputs and assumptions evaluation categories ranged between 2.36 for Accuracy, up to 2.68 for Comprehensiveness. The Comprehensiveness score was the highest across all PAs. Only 28 percent of the measures assessed by the evaluation team in 2015 used similar inputs, while 27 percent used similar assumptions. Both the Quality and the Comprehensiveness of the Inputs and Assumptions were found to be steadily increasing across the three program years, but the Accuracy rating dropped noticeably in PY2015.
- SCE's average ratings for the three inputs and assumptions evaluation categories ranged between 2.35 and 2.46. In 2015, forty-two percent of the measures used similar inputs by the evaluation team, while 44 percent used similar assumptions. The remaining measures used different or

adjusted inputs and assumptions. While SCE saw the highest results across all three metrics in PY2014, the PY2015 results were lower than the three-year average across all three metrics.

- SCG had the highest Quality of Documentation rating (2.67) and Accuracy rating (2.58) of the four PAs, however, only 28 percent of time did the evaluation us similar inputs and 34 percent of the time did the evaluation use similar assumptions. The remaining time different or adjusted inputs and assumptions were used by the evaluation team. SCG was also the only PA to see a significant increase in their inputs and assumptions ratings both over PY2014 results and over the three-year average.
- All SDG&E average ratings for the three inputs and assumptions metrics were found to be below 2.4. For all three metrics, SDG&E was found to have the lowest three-year average ratings. However, the evaluation team did use similar inputs for 36 percent of the measures, and similar assumptions for 44 percent of the measures.

FIGURE 6-15: COMPARISON MEAN INPUT AND ASSUMPTION RATINGS BY PA ACROSS PROGRAM YEARS



* Note that the scale has been adjusted to range between 1.5 and 2.9 so that the trend over the three program years can be more easily identified.

Similar to the Calculation Assessment Ratings, the overall result from this three-year comparison of inputs and assumptions sees some ups- and downs- in the model ratings, but no clear trend one way or the other across any PA (with the exception of SCG). There is no evidence of systematic changes made to address any inadequacies identified in the models. *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.*

6.3.8 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 6-10 presents these results.

TABLE 6-10: INCREMENTAL COST DOCUMENTATION BY PA

Parameter Examined	PA Incremental Cost Treatment			
	PG&E	SCE	SCG	SDG&E
Number of Measures Assessed* (n)	30	16	7	24
Number of Measures with Incremental Cost Populated (n)	23	15	7	11
Percent of Measures with Incremental Cost Populated	77%	94%	100%	46%

* 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.

SCE and SCG documented the majority of their incremental cost where appropriate, in 94 percent and 100 percent of cases, respectively. The other PAs should strive to report with equal accuracy. PG&E documented incremental cost in 77 percent of applicable situations, while SDG&E documented 46 percent of the time.

7 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 2014 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 2015, 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 internal PA processes, as well as coordinated activities across the PAs. However, while filing their Response to Recommendations (RTRs) for the 2014 report, the PAs did not explicitly agree to implement some recommendations; for example, the desire to balance appropriate level and duration for M&V with timely payment of incentives. Accordingly, this chapter identifies all unique findings and recommendations that have not appeared in previous evaluation reports by marking those paragraphs with a double asterisk (**).

The findings and recommendations in this report reflect CPUC policies and evaluation guidance applicable to the 2015 program year. The evaluation team is aware of relevant legislation, such as Assembly Bill 802 and Senate Bill 350, and associated CPUC Decisions that might change current policies.[85] Recommendations in this report have been made without any speculation surrounding future regulatory changes.

Findings and recommendations are organized into the following sections:

- 7.1 Gross Impact-Related Findings and Recommendations
- 7.2 Net-to-Gross-Related Findings and Recommendations

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

[85] This includes work-in-progress on the business plan framework for portfolio planning and the CPUC Decision and Resolution process.

As summarized in Table 7-1 (and Chapter 4), 2015 ex-post MMBtu lifecycle gross impact realization rates (LC GRRs) range by PA from 0.41 to 0.51.[86] Relative to 2014 custom impact evaluation results, these LC MMBtu GRRs *decreased* by about 25-30 percent for PG&E, SCE and SDG&E, and *increased* by four percent for SCG.

TABLE 7-1: 2010-2012, 2013, 2014 AND 2015 WEIGHTED PROJECT LIFECYCLE 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	2015 Mean Gross Realization Rate	2015 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	0.47	0.36 to 0.58
kW	0.46	0.35 to 0.58	0.44	0.28 to 0.61	0.74	0.34 to 1.14	0.50	0.34 to 0.67
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	0.41	0.34 to 0.49
kW	0.57	0.47 to 0.67	0.52	0.43 to 0.62	0.46	0.34 to 0.58	0.40	0.27 to 0.52
SDG&E 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	0.47	0.4 to 0.55
kW	0.82	0.46 to 1.17	0.76	0.57 to 0.95	0.63	0.54 to 0.71	0.73	0.48 to 0.99
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	0.51	0.38 to 0.64

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

Net-to-gross ratios (Table 7-2), ranging from 0.50 to 0.57, are similar in magnitude to previous evaluation results.[87] However, 2015 SCE NTG results are statistically significantly higher than 2014 results, although equal in value to 2013 results.

[86] 2015 ex-post gross impact results were also developed in this evaluation for MMBtu first year realization rates (FY GRRs), which range by PA from 0.51 to 0.55. Relative to the 2014 custom impact evaluation results FY MMBtu GRRs *decreased* for all PAs -- roughly nine percent for PG&E and SCG, 14 percent for SCE and 30 percent for SDG&E. 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 all PAs.

[87] NTG, as reported here, is inclusive only of free ridership effects (1-FR) and does not include spillover or market effects.

TABLE 7-2: COMPARISON OF 2010-12, AND 2013, 2014 AND 2015 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	2015 Mean NTGR	2015 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	0.53	0.49 to 0.58
SCE								
MMBtu*	0.49	0.47 to 0.50	0.57	0.52 to 0.61	0.46	0.42 to 0.49	0.57	0.53 to 0.62
SDG&E								
MMBtu*	0.48	0.46 to 0.50	0.59	0.55 to 0.64	0.51	0.47 to 0.56	0.50	0.44 to 0.57
SCG								
MMBtu*	0.49	0.40 to 0.58	0.66	0.59 to 0.73	0.62	0.60 to 0.65	0.57	0.54 to 0.61

* 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,[88] 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.
- To reduce continued moderate free ridership, PAs should consider changes to program features and implementation procedures designed to increase program influence.

[88] It is notable that the evaluation estimate of EUL differed from the PAs estimate 52 percent of the time. For those instances the evaluation-derived average EUL was smaller than the ex-ante average EUL by roughly 3 years, representing a 26 percent reduction in the ex-ante EUL claim for that subset of observations. It should be noted, however, that result varied substantially by PA. As noted in Chapter 4 LC GRR results are lower than FY GRR results and this EUL difference is *a key factor* driving down the LC GRR results.

7.1 GROSS IMPACT-RELATED FINDINGS AND RECOMMENDATIONS

As presented in Chapter 4, calculation methods, inappropriate baselines, ineligible measures and operating conditions 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 Calculation Methods
- Project Baseline Specification
- Project Operating Conditions
- The State of Ex-Ante M&V

7.1.1 Underperforming Projects

All PAs had projects with negative and/or zero GRRs, and these served to lower the weighted realization rate considerably. Out of 148 M&V points, 30 projects, or 20 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 4, and 22 of the cases were due principally to one of two factors – inappropriate baseline or ineligible measures. For each of the PAs, these two issues had a substantial downward effect on the resulting ex-post lifecycle savings estimates, ranging from a 20 to more than 40 percent reduction, and negated some of the largest project-level claims. Other factors that had a large downward influence on individual project-level savings estimates includes calculation methods, inoperable measures and operating conditions.

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 lessons learned from 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:***

- ***PAs should improve 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 4: 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 PAs should 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. For example, projects involving fuel switching must pass the 3-prong test or are otherwise ineligible.
- PAs should screen measures for eligibility, including removal of maintenance measures and assurance that projects meet program eligibility performance thresholds.
- 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, ***the PAs should carefully review each of the 30 FSRs listed in Section 4.4.2, Table 4-6, 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, baseline specifications that do not meet post-installation service requirements and conditions, calculations that include errors, lack of validation of equipment specifications and modeled performance, and failure to apply the non-regressive baseline rule.
- The PAs should 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 resulting lifecycle ex-post gross savings estimates.

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, the PAs need to thoroughly document 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 improved.***
- **Recommendation:** ***To improve project eligibility screening the PAs should 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 should examine Appendix F, which includes a list of every project where the evaluation overturned the PA specified project type or baseline type.

- ****Recommendation*: PAs should push participating customers to higher levels of efficiency in order to build in a savings buffer above ISP/code/non-regressive baselines and thereby have greater assurance of project eligibility and achievement of ex-ante saving claims.***

7.1.2 Project Calculation Methods

Recommendations to improve calculation methods and protocols are presented in this section. As noted in Chapter 6 and Section 4.5, the ex-ante calculations for an array of projects were lacking in terms of the calculation method applied and incorporation of correct inputs that describe typical or representative operating conditions. Improvements to capturing operating conditions and enhancing associated model accuracy are discussed in both this section and Section 7.1.4.

As discussed in Section 4.4 downward adjustments to ex-ante first year claims due to calculation methods was the most important discrepancy factor in the M&V sample. Calculation method issues was the leading downward factor for SCE and ranked as the number two factor for both PG&E and SCG.

Finding: Impact Methods and Models

For the majority of projects included in the evaluation gross impact sample the ex-post evaluation used a different model or adjusted the PA ex-ante model. Furthermore, the evaluators used different inputs and assumptions for the majority of projects in the sample. 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: *PAs should 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 needing improvements to 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. CPUC guidelines should be followed, including the estimation of savings when non-IOU supplied energy sources are used, such as performing hourly net grid impact analysis. 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. Calculations should be developed using proven tools.
- Further, the PAs should include in each application file the live, unlocked, non-password protected spreadsheet models. The PAs should ensure the final model is stored in each file and record key model inputs and outputs, documented using data or observed conditions.
- **Recommendation*: PAs should 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: *PAs should conduct periodic due diligence to ensure programs adhere to PA and CPUC impact estimation policies, guidelines, and best practices*.** 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: *The PAs should prioritize M&V reviews for <u>all</u> large projects.*** Based on the distribution of custom projects by size observed in 2015 a census of large projects in strata 1-3 ranges by PA from just a handful or projects to less than 50, and represents roughly 40 to 60 percent of ex-ante savings claims. The purpose would be to ensure that CPUC M&V standards are being met for the treatment and documentation of program ex-ante savings. This would reduce risk to ex-ante claims, and should focus on proper baseline documentation, appropriate eligibility screening, CPUC-approved M&V planning and implementation, and the development of robust and accurate savings estimation models and results.
- ****Recommendation: *For certain applications, such as where the baseline is represented by the pre-existing equipment and pre- to post-installation conditions are stable, PA use of an IPMVP Option B or C regression model may be preferable to other calculation-based approaches.*** Regression models should also account for all non-routine adjustments, as facilities often undergo changes unrelated to program efficiency-based improvements, and savings estimates should be normalized for production and weather differences. It is also critical that the measure-impacted accounts be properly identified and used in regression models. Regressions may serve to better

bound the savings and may also be used as a sanity check of results derived using other calculation approaches.

- **Regression models should be informed by longer duration trend data whenever feasible.
- **For regression models involving both energy consumption data and production data (i.e., energy intensity), a variety of models should be attempted using differing time intervals, such as daily versus hourly, in order to identify model-based estimates with the best fit regression curve.
- **Where regression models are used the R squared values should be 0.70 or higher and the CV(RMSE) values should be lower than 15 to 20%.

- ****Recommendation:** ***For NRNC whole-building projects the PAs should use the non-compliance mode to estimate savings and compliance mode to demonstrate project eligibility.***
- ****Recommendation:** ***The PAs should review all modeling weaknesses and areas for improvement noted in Section 4.5*****.**

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, production records 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 conditions, and to calibrate models.
    - **Calculated savings should be based on robust data sets representing longer-term and stable operation of equipment and systems. PAs should collect appropriate trend data that demonstrate typical operation, and ensure that M&V data used to estimate ex-ante savings estimates properly account for variation in weather, seasonality, equipment performance and production schedules/operations. Where variability is present, PAs should wait to claim savings until a more confident savings estimate, based on typical operation, has been developed.
    - **For pump efficiency improvement projects, historical energy usage and production data should be used to derive estimates of kWh/acre-foot and OPE.

- **PAs should encourage participating customers to collect and retain data for purposes of conducting project-level M&V, especially where instrumentation is available.
- **In the absence of trend data PAs should alternatively use manufacturer equipment specifications to inform calculation inputs.
- **Where M&V data collection is infeasible or impractical, inputs and assumptions should be based on conservative assumptions.
- **PA models should use custom rather than deemed variables in calculations where inconsistencies exist between project conditions and assumptions that define the deemed calculation approach.

- **Recommendation:** ***Regarding peak demand analysis, adopt CPUC protocols and procedures as they relate to the DEER-based California climate zone peak period definition.***[89] Peak impact estimates should reflect loads during the California climate zone three-day period. 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.

7.1.3 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 for PG&E, representing the largest discrepancy factor for that PA, and was among the top four discrepancy factors for all the PAs besides SCE.

Finding: PA Baseline Changed by Evaluation

There was generally good agreement on project baseline when comparing PA and evaluator selections (72 percent agreement across all PAs and projects). However, there was less agreement surrounding project type designations (58 percent agreement), which should be used as a determining factor for

[89] From the CPUC Energy Policy Manual, version 5: “The definition of peak megawatt load reduction contained in the most recently adopted DEER shall be used to estimate and verify peak demand savings values. The DEER method utilizes an estimated average grid level impact for a measure between 2 p.m. and 5 p.m. during a “heat wave” defined by three consecutive weekdays for weather conditions that are expected to produce a regional grid peak event.”

proper baseline selection. Add-on, new construction and ROB projects were the most commonly overturned project types across all PAs, followed by ER.

- **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 measures 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), as well as properly validating that installed measures do not entail like-for-like replacements from an efficiency perspective. If the efficiency of the pre-existing equipment is higher than the otherwise accepted 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. The PAs should 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.
- **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. 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 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. Furthermore, PAs should conduct independent research for the purposes of identifying project-level ISP baseline and provide a comprehensive narrative backed up by data that correctly identifies ISP.

- ****Recommendation**: ***Appropriate interpretation and application of code requirements is needed, including the need to consider and possibly examine a broad array of codes and requirements that may be relevant for a given project.*** During the last decade of evaluations in California, baselines have been defined using local codes, regional codes, state codes and federal codes, spanning energy-based requirements, safety requirements, and air or water/wastewater quality requirements, as well as facility service and functionality requirements. During application review the PAs should carefully consider all relevant code requirements and update ISP and other baseline determinations for relevant measures.

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.
- ****Recommendation**: ***PAs should demonstrate the availability of selected baseline equipment when establishing ISP. Ordinarily this would include obtaining quotes for available new, less efficient, but functionally equivalent equipment (baseline).*** 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. PAs should demonstrate that baseline equipment selected represent a feasible option, given facility constraints and production needs.
- ****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, select a representative baseline period, 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.

7.1.4 Project Operating Conditions

The operating conditions discrepancy factor is the 4th largest 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 inform 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. For SDG&E operating conditions accounted for about one-third of all downward adjustments to ex-ante claims, but was less important for the other PAs.

- **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. This includes use of trend data to generate performance curves and estimate power consumption. Also, assumptions used should reflect conservative values supported by strong evidence from secondary sources.

  PAs should increase the use and improve incorporation of, data collection and monitoring to ensure a meaningful and accurate set of inputs or assumptions surrounding operations. Post-retrofit inspections should fully incorporate verification of measures, proper installation and operation, and any observed or otherwise known changes or deficiencies. PA staff should check that pre-installation and post-installation reports are well organized and complete, with measure counts, changes in operation, efficiency values, and operating parameters.

- The PAs should 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. For example, operating hours used in calculations should reflect observed conditions via verification and M&V. Additional due diligence in this area is needed when loads are variable, including projects with seasonal variation in production and operations. 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 observed 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.

  **Measures such as agricultural pumps require lengthier trend data sources, given that operations can be greatly affected by weather, including drought conditions, and water availability.

- For projects entailing the use of simulation models, models should be re-run after the equipment is commissioned and building loads represent steady state operation.

  **For new construction projects associated with either tenant improvements or new buildings, PAs should wait to file claims once the project is fully built out and occupied. A certificate of occupancy can be used to inform the timing of claims. CPUC evaluation guidance is to model savings based on the as-found conditions.

- 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.

7.1.5 The State of Ex-Ante M&V

Both the Chapter 4 gross impact and Chapter 6 PPA results, including trends from recent evaluations, generally do not point to PA improvement. Project ex-ante treatment shows a lack of attention to CPUC guidance, decisions, previous evaluation results, ex-ante review-based directives, and adequate use of documentation and data-derived calculation methods and inputs. Even some of the largest projects demonstrate a lack of due diligence.

Finding: PA M&V Improvement is Needed

Recommendation: ***It is recommended that a statewide document, similar to the PPA form, be developed for use by all PAs for custom claims.*** The project practices assessment (PPA)[90] 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 perhaps in a less systematic way, and improvements to forms and processes should have a positive outcome on results. In addition to the form itself, Appendix E provides detailed descriptions of PPA scoring criteria that will help PAs ensure they are adequately capturing and documenting the relevant information. 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.

[90] Project Practices Assessment reviews were conducted for all completed measurement and verification (M&V) sample 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.

The 2013 through 2015 PPA results, combined with GRR and NTGR findings, provide a solid baseline from which to continue tracking PA performance. Given that ex-ante review process began in earnest in January of 2012, effective PA processes and procedures and related improvement is overdue.

7.2 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 5 of this report.

Finding: Free Ridership for Custom Projects Remains Elevated

On a statewide basis, the NTGR averaged 0.55. This demonstrates a moderate increase from the PY2014 NTGR of 0.51, and NTGR results indicate a medium[91] level of free ridership and a resulting medium level of program influence. Note that this value continues to be similar in magnitude to NTGRs from the past several evaluation cycles, as shown in Table 7-3. 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.

TABLE 7-3: STATEWIDE INDUSTRIAL AND CUSTOM PROGRAM[92] EVALUATION NET TO GROSS RATIOS, PROGRAM YEARS 1998-2015

(1 – Free Ridership)	1998	1999	2000	2001	2002	2004-2005	PY2006-2008		2010-2012	2013	2014	2015
							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	Statewide MMBtu - 0.55

*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. There were also instances where incentives

[91] 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).

[92] 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.

were provided to firms that were already very advanced in their adoptions of energy efficiency, such as water/wastewater plants, and companies with established energy efficiency procurement policies or mandates, including national chain and big box stores.

- **Recommendation:** ***Adopt procedures to identify and affect projects with low program influence.*** The PAs should carefully review projects during the project development stage for potential issues associated with a high likelihood of very low program influence. This process should provide timely feedback to program implementers regarding the estimated level of program influence. This would afford implementers an opportunity to influence projects found to have low program attribution by encouraging project decision makers to adjust the project scope to higher efficiency levels, where warranted.
- **Recommendation:** ***Adjust the set of technologies that are eligible for incentives.*** Periodically review the list of qualifying measures for each program and eliminate eligibility for those that have become standard practice. At a minimum, such reviews should take place annually. Measures that are already likely or very likely to be typically installed should not qualify for incentives. Although identification of such measures can be difficult in practice in the industrial sector, a number of such measures can be identified through investigation of industry practices (for example, interviews with manufacturers, distributors, retailers, and designers), analysis of sales data, and review of evaluation results. In determining which measures to retain and which to eliminate, a balance must be struck between reducing free ridership and avoiding significant lost opportunities. Ideally, sub-technology niche markets can be selected for the program that are less well established, but where substantial technical potential still lies.

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

  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.