# Evaluation and Regulatory Teamwork: Closing the Custom M&V Gap

Kris Bradley, Itron, Inc., Oakland, CA Kay Hardy, California Public Utility Commission, San Francisco, CA Peter Lai, California Public Utility Commission, Los Angeles, CA

## ABSTRACT

This paper examines an on-going multi-year improvement process for custom energy efficiency projects and programs targeted at the non-residential sector. The process uses a combination of policy guidelines, ex-ante review, program requirements, ex-post evaluation, and QA/QC procedures to improve both custom impact estimates and custom incentive programs. Non-residential, custom programs present unique opportunities for large, cost effective energy efficiency impacts. At the same time, custom programs often face difficult challenges, both in terms of effective program implementation and accurate estimation of impacts. These challenges are due, in part, to the complexity and diversity of the associated projects, as well as the difficulty of aligning procedures, policies, and quality control to an element of the program portfolio that is characterized by individual project interactions rather than standardized transactions. These issues have manifested in lower than expected ex-post gross impact realization rate results for custom projects and programs.

The regulatory and program administrator processes described in this paper were developed to improve the accuracy of ex-ante savings estimates through a combination of proactive regulatory activities and policies and innovative evaluation approaches and products. The paper features an extensive effort to quantify and remediate differences in regulator and administrator approaches and perspectives on key elements of impact estimation, including baseline determination, project eligibility screening, savings calculation methods, and modeling of observed operating conditions, among other elements.

#### Introduction

This paper addresses incentive programs in California that focus on custom offerings. These programs have traditionally been implemented by the investor-owned utilities, Pacific Gas and Electric, Southern California Edison, Southern California Gas Company and San Diego Gas and Electric. These and other entities that implement programs in California are referred to as PAs or "program administrators." The programs included in this custom impact evaluation address industrial and manufacturing facilities, including water supply, water treatment and wastewater treatment, oil and gas extraction, and oil refining and production. The evaluation also addresses large commercial and agricultural custom program offerings.

Evaluation-based gross impact realization rate results for these custom offerings in California have consistently been lower than desired. This could be due to differences in ex-post as compared to ex-ante operating conditions or to a broader disconnect between program (ex-ante) project treatment and evaluation (ex-post) treatment for a number of important project-level, impact estimation elements. Analysis of the reasons for the observed gap have pointed to a broad set of differences in ex-ante and ex-post assessments, rather than to solely ex-ante forecasting errors associated with post-installation operating conditions. To address this gap, a two-prong approach has been developed and implemented, with a focus on 1) regulatory feedback and 2) evaluation feedback. This approach is illustrated in Figure 1.



Figure 1. Combined Regulatory and Evaluation Feedback System

While the regulatory feedback from the California Public Utility Commission (CPUC) includes the provision of policy guidance and associated program requirements, the anchor to this effort consists of a relatively new ex-ante review (EAR) activity. EAR involves the parallel participation by CPUC staff and their contractors in the development of ex-ante savings estimates for a sample of selected projects. CPUC staff become involved through the EAR process at an appropriate time falling between the project concept stage and the finalization of ex-ante savings estimates. Here CPUC staff and their contractors coach and instruct program administrators in proper development of savings estimates and establish enhanced policy guidance and associated program requirements. This process serves to narrow the gap between evaluation and program treatment, both broadly and under specific circumstances and conditions, even for certain projects/measures. This new EAR activity has been in place and operating across all program administrators since January of 2012, following the establishment of the associated architecture for this activity during the summer of 2011 in CPUC Decision 11-07-030. The establishment of ex-ante review (EAR) is discussed in CPUC Decision 09-09-047,<sup>1</sup> which requires that CPUC staff review and approve ex-ante impact estimation approaches and ex-ante savings for custom measures. The ex-ante review process is designed to provide constructive early feedback to program administrators 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 surrounding impact estimation methods and documentation.

The evaluation report and associated dialogue and feedback mechanisms are also designed to influence program impact results. Recent reports have focused more specifically, project-by-project and overall, on discrepancy factors and the differences that are found to exist between evaluation-supported conclusions and findings versus the program claims. By exposing discrepancy factors and differences it becomes possible to correct within-program processes and procedures, and thereby narrow the savings estimate gap.

These regulatory and evaluation activities support one another using information sharing and crossactivity feedback mechanisms. This ensures that the same rules are enforced and approaches are encouraged across the two-prong program improvement process. Importantly, when new regulatory guidance and requirements are put in place, often based on EAR-derived conclusions, the evaluations and programs must subsequently adopt this new precedent. The provision of feedback and associated information sharing is designed to support this objective.

<sup>&</sup>lt;sup>1</sup> The decision may be found at the following web link: <u>http://docs.cpuc.ca.gov/published/FINAL\_DECISION/139858.htm</u> 2015 International Energy Program Evaluation Conference, Long Beach

# **Historic Impact Evaluation Results and Findings**

Custom program evaluation results in California have historically yielded lower than expected expost gross impact realization rate results. This is demonstrated in Table 1, which presents results from the last three consecutive evaluations completed – the PY2006-2008,<sup>2</sup> PY2010-2012<sup>3</sup> and PY2013<sup>4</sup> evaluation reports. The first-year gross impact realization rate (GRR) results, which represent the ex-post evaluation results divided by ex-ante savings estimates, generally range, on average, from 0.60 to 0.70. The results indicate that there are substantial differences between ex-post and ex-ante estimates and that additional effort is warranted to bridge this gap.

Program	2006-2008 FY GRR Results			2010-2012 FY GRR Results			2013 Draft FY GRR Results		
Administrator	kWh	kW	Therms	kWh	kW	Therms	kWh	kW	Therms
PG&E	0.49	0.46	0.68	0.68	0.58	0.7	0.74	0.53	0.74
SCE	0.72	0.65		0.60	0.61		0.54	0.64	0.54
SCG						0.64			0.69
SDG&E				0.46	0.84	0.64	0.75	1.03	0.75

Table 1.	Comparison	of California	Custom Program	Evaluation	Results
I abic Ii	Comparison	or cumorniu	Custom rogram	L'uluulon	results

Examination of the results indicates a general upward trend in GRR, with the exception of one administrator. This might be an indicator of improvement in ex-ante savings estimation during this period. However, the reader is also cautioned that market conditions may also be a contributing factor and some of the differences are not statistically significant. The PY2006-2008 results in particular were affected by the economic downturn/recession at the time of evaluation. The evaluators noted that many participating businesses were affected by facility closures and other downsizing factors that also resulted in lower GRR results. During the PY2010-PY2013 period these effects lessened, and might also explain the upward (improvement) trend in GRR. Stating this another way, evaluation GRR results in PY2006-2008, as well as PY2010-2012, may have been artificially low due to effects of the economic downturn, and therefore the trend of improvement in GRR may also be artificial.

During the course of these evaluations the lower than desired GRR results have been explained, in part, as a function of program practices and issues. For example, problems associated with program (exante) baseline selection, or the need for estimates to be based on more conservative assumptions, more frequent use and application of M&V, and more careful scrutiny and documentation of projects, including pre-conditions, among others. These and several other examples not listed were all present and identified during the PY2006-2008 evaluation, and still persisted to some degree in PY2013 despite considerable efforts to rectify those issues and provide feedback to improve ex-ante savings estimates. It was common in PY2013 for ex-ante baseline to be incorrectly specified; for example, the energy efficiency project receiving an incentive was sometimes the only feasible technical solution for the customer, which yielded no savings for some projects. That said, there is improvement to report, based on collaboration among evaluators, regulators and program administrators.

As a result of the persistence of issues and modest gains to date in GRR results, regulatory feedback and evaluation feedback have both adopted a keen focus on improvement, and greater specificity in the

<sup>&</sup>lt;sup>2</sup> <u>http://calmac.org/publications/PG%26E\_Fab\_06-08\_Eval\_Final\_Report.pdf</u>

and http://calmac.org/publications/SCIA\_06-08\_Eval\_Final\_Report.pdf

<sup>&</sup>lt;sup>3</sup> http://calmac.org/publications/2010-12\_WO033\_Custom\_Impact\_Eval\_Report\_Final.pdf

<sup>&</sup>lt;sup>4</sup> http://www.energydataweb.com/cpuc/deliverableView.aspx?did=1265&uid=0&tid=0&cid=

<sup>2015</sup> International Energy Program Evaluation Conference, Long Beach

evaluation-based analysis and reporting procedures in an effort to yield actionable findings and recommendations.

### **Evaluation Feedback**

The CPUC began managing impact evaluation efforts in PY2006. The program portfolio had expanded tremendously at that time in response to increased goals. Previous evaluation efforts in California had been managed by the program administrators. A more rigorous custom impact evaluation approach was implemented starting in PY2006. Greater rigor was enabled by greater expenditures in general and for M&V per project sampled. Greater resources enabled increased focus on issues that were challenging to address and often required extra time for analysis, for example, assessing whether early retirement projects demonstrated program influence, estimating life cycle-based savings estimates, using more intensive and longer term sub-metering, expanding analysis of very large and complex projects involving multiple fuels and end uses, ensuring projects met program and policy eligibility requirements, and so forth.

In addition, custom evaluations began at that time to focus on specific ex-ante project treatment issues and actionable evaluation-based recommendations to address ex-ante deficiencies. In subsequent evaluations during PY2010-12<sup>5</sup> and, more recently, PY2013,<sup>6</sup> this has become an even greater focus. In PY2010-12 very specific findings and associated recommendations were provided. For example, these evaluations included examination of discrepancy factors that describe the reasons behind differences in exante and ex-post savings estimates and quantify the total effect of such factors on realized savings. Most important among those factors is ex-ante treatment of baseline selection, operating conditions, calculation methods and eligibility determination. Other findings that emerged or were highlighted in the PY2010-12 evaluation included insufficient quality of PA documentation and tracking data, lack of sufficient quality control for gross impact savings estimates, and the need to more carefully screen for ineligible projects.

The PY2013 custom evaluation has continued to emphasize identification of ex-ante savings estimation issues and associated discrepancy factor analysis, as well as actionable recommendations to address those items. Selected PY2013 results are presented in Figures 2 through 4 to identify factors that lead to a reduction in ex-post savings estimates relative to ex-ante, including the frequency with which discrepancy factors occur, as well as the respective evaluation-based reduction to ex-ante claims. Figure 2 presents M&V results for the evaluation sample, segmented by tracking system records where the ex-post evaluation savings were – reduced, increased or equal – relative to ex-ante claims.

<sup>&</sup>lt;sup>5</sup> http://calmac.org/publications/2010-12\_WO033\_Custom\_Impact\_Eval\_Report\_Final.pdf

<sup>&</sup>lt;sup>6</sup> http://www.energydataweb.com/cpuc/deliverableView.aspx?did=1265&uid=0&tid=0&cid=



Figure 2. PY2013 CA PA Custom Impact Results Segmented by Evaluation Adjustment

Ex-ante savings were more likely to be overstated than understated. 157 out of 240 records in the evaluation M&V sample had a downward effect on the average unweighted GRR. The GRR for that grouping is 0.51, indicating that ex-post evaluation realized savings were 51 percent of claimed savings. A much smaller number of records, 41, experienced an upward effect on savings claims, and had a smaller average and total effect on the overall GRR result. A similar number of sample points, 42, had an evaluation result that was equivalent to the ex-ante claim (i.e., GRR = 1.00).

To address the factors contributing to the downward effect on ex-ante impact claims, results were developed by discrepancy factor category. Across all PAs the factors that had the largest downward effect on impact claims are presented in Figure 3 and include inappropriate baseline, operating conditions, calculation methods and measure eligibility.



Figure 3. Discrepancy Factors with the Largest Downward Effect on Impact Claims

Inappropriate baseline explains the downward effect on impact claims for 30 tracking system records, and was the factor that accounted for the largest reduction in claimed savings (as indicated by the *difference* between the ex post and ex ante bar heights in the figure), with a 0.4 million MMBtu reduction and an unweighted gross impact realization rate (GRR) of 0.21 for that grouping of sampled records. Also notable is that 14 of those 30 records resulted in a GRR of zero, where the program installed equipment were equivalent in efficiency to baseline equipment, most often based on either industry standard practice or code-based baseline determinations.

Operating conditions was the factor that accounted for the second largest reduction in claimed savings, occurring in 48 of the sampled tracking system records, with an unweighted GRR of 0.71. Operating conditions address differences in ex-post conclusions surrounding production levels, operating hours, equipment load profiles and equipment sequencing, among others.

Calculation methods and ineligible measures accounted for the third and fourth largest reduction in impact claims. Calculation method was the most frequently cited reason for reduction and ineligible measures resulted in the lowest GRR of all factors at 0.06.

Given the magnitude of the associated savings reduction and the frequency with which these factors were observed in the M&V sample, each of these four factors represents critical issues that have been prioritized for correction by PA QA/QC procedures. To help address this pressing need, the evaluation developed sub-categories that better explain the circumstances that led to discrepancy factor reductions. These subcategories are illustrated for one PA, PG&E, in Figure 4.<sup>7</sup> For example, for the operating conditions discrepancy factor, 19.4 percent of the observations were associated with differences in estimated operating hours.

<sup>&</sup>lt;sup>7</sup> http://www.energydataweb.com/cpuc/deliverableView.aspx?did=1265&uid=0&tid=0&cid=



Figure 4. Most Influential Discrepancy Factors that Caused Downward Adjustments for PG&E

Some new findings also emerged in PY2013 surrounding differences in evaluation estimates for effective useful life values (EULs). There were differences in evaluation versus program EULs in 46 percent of all evaluated tracking records, which led to an average reduction in evaluation EUL by 2 years. This had a substantial downward effect on GRR results. Evaluation sources for EUL principally differed from program sources for two reasons. First, in the case of add-on measures, the evaluation treatment of EUL for the new equipment was sometimes capped based on the remaining useful life (RUL) of the pre-existing or host equipment. For example, for a new VSD on a pump, the EUL of the VSD might be capped based on the RUL of the host pump equipment unless it can be demonstrated that the VSD will continue to operate beyond the time of expected pump failure. Second, different approaches were applied in extracting Database for Energy Efficient Resources (DEER)<sup>8</sup> database EUL values.

The PY2013 evaluation also included a new analytic framework – project practices assessment or PPA. PPAs examine custom project impact estimation methods and procedures, and facilitate an assessment of PA ex-ante impact estimation procedures and documentation for custom projects. These reviews feature assessments of project compliance with ex-ante review guidance and requirements, and conformance with policy guidance, with an emphasis on ex-ante gross savings development and methods. This PPA framework also quantifies differences of opinion between the evaluation perspective and that of the program's. PPA emphasizes eligibility treatment, baseline determination, project cost treatment, RUL and EUL determination, and gross impact calculation methods, inputs and assumptions. PPA builds off of earlier work in PY2010-12, which was referred to as the lower rigor assessment (LRA), and provides for quantitative results in these specific areas that can then be utilized to identify program issues and support recommendations for improvements in ex-ante savings estimation, as well as other areas of assessment and documentation.

<sup>&</sup>lt;sup>8</sup> http://www.energy.ca.gov/deer/

<sup>2015</sup> International Energy Program Evaluation Conference, Long Beach

PPA results provide a few useful statistics to the PAs on the frequency of differences between expost and ex-ante conclusions. For example, the PPA assessment shows that across PAs the evaluation and program agree on project-level baseline labels over 70 percent of the time – such as existing equipment, Title 24, Title 20, industry standard practice, federal regulations, etc. For those cases where there are differences in baseline assessment, the results identify what the over-turned ex-ante baseline was and what the evaluation determined it to be. The most frequently observed cases where the ex-ante existing equipment baseline was over-turned was in instances where the evaluation concluded that baseline is industry standard practice. Another result involves the frequency with which the evaluation used the same model to estimate savings – 46 percent of observations. For the remaining observations the evaluation used a different model or adjusted the ex-ante model.

Similar to the discrepancy factor analysis described above, the intent behind the PPA results is to point specifically to the areas where improvement in ex-ante savings estimation procedures and documentation is needed. It is hypothesized that correction to identified areas will improve ex-ante savings estimates. Also, PPA facilitates a longitudinal assessment of PA performance over time, similar to Table 1 above. For example, given that ex-ante and evaluation defined EULs were found in PY2013 to be different 46 percent of the time, PPA facilitates an ongoing assessment of that same comparison, and progress towards the desired improvement.

#### **Regulatory Feedback**

In tandem with the CPUC beginning to manage evaluations starting in PY2006, the CPUC also took a more active role in program regulatory processes. This includes the review and update of work papers associated with deemed measure savings claims, enforcement of Database for Energy Efficiency Resources (DEER) savings estimates and approaches/inputs, greater participation in working groups with the PAs, and more intensive work to refine policy/guidance and program requirements. However, the focus of this section of the paper is on the development and implementation of EAR processes and procedures. EAR is an evaluation-oriented regulatory approach that could potentially be applied in other jurisdictions where custom energy efficiency programs operate, in an effort to improve program processes and procedures and evaluation results.

The activities under EAR include gross impact engineering review for energy efficiency incentive applications and projects being implemented by PAs prior to final approval of those projects. The establishment of ex ante review is discussed in the CPUC Decision on the petition to modify Decision 09-09-047,<sup>9</sup> which requires the Energy Division (ED) to review and approve ex-ante impact estimation approaches and ex-ante saving estimates for non-DEER ("custom") measures.

EAR activities initially emphasized project review for selected project applications prior to finalization of ex-ante saving estimates/claims and the associated incentive payment. The CPUC staff-appointed teams work with each PA to improve application-based estimates of savings using appropriate rigor. These teams communicate using both written correspondence and weekly meetings to synthesize the regulatory position on an array of related topics, including assessments of baseline selection and related project cost documentation, measure eligibility, conformance with program rules, conformance with CPUC policy and guidance, and thoroughness, accuracy and appropriateness of engineering models being applied and all inputs and assumptions being used. Unlike evaluation sampling, these projects are selected based upon a wide array of considerations, but often based on the likelihood of important evaluation issues being present that have not yet been adequately addressed by the programs. For example, past efforts have shown that larger projects don't always receive the level of scrutiny that is warranted, and that certain technologies are likely to have common problems that might be related to eligibility, baseline selection, fuel switching considerations, and so forth.

 <sup>&</sup>lt;sup>9</sup> The decision may be found at the following web link: http://docs.cpuc.ca.gov/published/FINAL\_DECISION/139858.htm
2015 International Energy Program Evaluation Conference, Long Beach

During the review and dialogue on selected projects, the EAR team sets expectations for projects and provides instruction regarding proper protocols and procedures for M&V. Where is measurement appropriate? What level of rigor is expected? What methods are acceptable? Where codes and standards do not apply, what defines industry standard practice? The EAR process began with significant levels of demonstration by the CPUC staff-appointed EAR teams (EAR teams) – conducting on-site inspections, running independent calculations, researching industry standard practices, and writing comprehensive reports on each project outcome, often referred to as a disposition. However, over time the EAR process has evolved.

During the second phase, roughly one year after the EAR process had operated in earnest, the EAR teams began to emphasize provision of guidance for selected projects in an effort to build enhanced internal PA QA/QC of custom applications and reduced reliance on the EAR team in conducting inspections, carrying out calculations and conducting research. Use of comprehensive reports persisted.

During the third phase, the PA and CPUC collaboration began to more comprehensively address information dissemination within the PAs to bring improved methods and rigor to all projects, not just those selected for EAR. One worry was that during the first two phases the PAs were very focused on the appropriate treatment for selected EAR projects, but that old habits and methods persisted for the remainder of the project applications. The PAs have developed tools and training in an effort to improve internal processes and procedures. This is still a work in progress, as demonstrated by the PY2013 gross impact results and PPA findings.<sup>10</sup> More focus in this area is needed for some of the PAs and likely additional resources to facilitate process improvements.

A newer fourth phase is also emerging in which the EAR teams are also addressing tracking systembased claims in order to verify that approved CPUC methods and policy are applied more broadly to the thousands of custom tracking system claims each year. For example, if a particular EAR project finds that injection molding machine baseline has shifted to electronic machines, then injection molding machine projects that come after the date of that particular disposition can be scoured to verify appropriate use of baseline. In addition to checking for conformance with general dispositions, tracking system records are also checked for eligibility considerations, program rules and policy considerations.

The EAR process to date has been successful as an instructional tool. By leading the PAs through an array of selected projects, the PAs are learning and achieving greater independence in developing CPUC-approved impact estimates for custom projects. Custom projects are complex and often unique, occurring in customer facilities ranging from laboratories to oil well fields, and must address issues that include market considerations, code requirements such as air quality considerations, varying levels of customer sophistication, detailed hourly simulation models, model calibration, model normalization and many other complex factors. As such it takes time for EAR processes to achieve results and being patient and persistent is necessary to achieve the desired improvements.

EAR also works well in concert with traditional evaluation. Evaluation is a crucial instrument for measuring success in the transformation of ex-ante impact estimation. Evaluation uses random sampling to examine ex-ante accuracy for a representative sample of projects.

It has been very challenging to affect all projects through information dissemination within the PA program systems. The larger PAs work with both internal staff and external third-party implementers. The scale of dedicated staff required to complete thousands of projects per year is quite large. Transforming the methods and habits that were previously formed for such a large number of program staff has not yet been successful. To some extent the effectiveness of the EAR effort has been hampered by the grand scale of custom programs in California. Expanded PA staff resources to address this situation should be considered.

<sup>&</sup>lt;sup>10</sup> It is important to recognize and acknowledge that it can take many years for changes in ex ante impact estimation practices to manifest in completed program projects and ex post evaluation given the sometimes lengthy custom project development cycles that can often last many years.

## **Coordination Between EAR and Evaluation**

One of the more challenging aspects to the two-prong EAR and evaluation feedback enterprise is ensuring that the evaluation stays up-to-date with policy and guidance being set by the EAR process and that the EAR process stays informed when evaluation precedent is set on a particular topic or situation. There are two distinct teams operating in parallel, both working with the PAs. Contradiction in opinion or project treatment might compromise both processes. A uniform message is needed to maximize improvement and not undermine the improvement process.

Given this architecture, three-way communication is critical. Some solutions involve tracking, posting and distributing information when a precedent is set. The EAR and evaluation activities support one another while working in parallel. Formal dispositions are developed by the EAR team and subsequently used to inform evaluation decision making. Likewise, when the evaluation work uncovers ex-ante impact estimation issues or program/measure problems, the EAR team can formalize treatment, using dispositions or other regulatory channels, or use evaluation findings to inform sampling of selected projects to further explore the issue at hand.

Evaluation gross impact (GRR) results and PPA findings provide the ultimate proving ground for how well the programs are functioning and the extent of any improvement. An important difference between EAR activities and evaluation activities is that evaluation activities involve random sampling of projects, a very small percentage of which are projects formerly selected for review under EAR. EAR sampling addresses projects that represent just a few percentage points of the total. The evaluation examines mostly projects that were never selected for EAR treatment, and therefore an opportunity to examine whether or not EAR activities are leading to change for the broader set of projects in the programs, or if its effects are largely constrained within the EAR sample. The intent is to have the EAR activities transfer to the full program population of projects and the evaluation provides a unique opportunity to measure and quantify the extent to which that objective is observed.

### **Program Administrator Incentive Mechanism**

An incentive mechanism for the PAs, designed to encourage PA implementation of energy efficiency, was established in California for the PY2013-14 program cycle. Various incentive strategies have been used in California since the 1990's, in response to the aggressive energy efficiency goals that are often sought in the state. The latest incarnation for the PY2013-14 programs is the Efficiency Savings and Performance Incentive Mechanism, or ESPI.<sup>11</sup> The maximum potential payout to the PAs through ESPI is \$126 million. But consistent with the name of this incentive mechanism, earnings through ESPI are dependent upon savings achieved and "performance."

Due to the importance of custom programs in terms of the contribution to overall portfolio savings claims, custom programs represent a considerable share of the incentive levels that can potentially be paid out to each PA. The custom portion of the incentive is based on both an assessment of custom program savings and performance. The savings assessment for custom programs is based on ex-post evaluation impact results for electric energy, electric demand and natural gas savings. Refer to Table 1 above for the GRR-equivalent of savings achieved. The performance assessment is based on metrics that are qualitatively scored by the CPUC, and are generally designed to assess administrator engagement, quality, responsiveness and improvement in the EAR process.

The ESPI payment metrics and approach is complex and it is not the intent of this paper to provide a detailed explanation of the ESPI approach or the payouts received. Nonetheless, it is important to note that ESPI provides encouragement to the PAs to improve ex-ante savings estimation and increase savings – the focus also of both the EAR regulatory effort and recent evaluations of custom programs in California.

<sup>&</sup>lt;sup>11</sup> http://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M076/K775/76775903.PDF

The ex-post evaluation component of the incentive reward is tied to achieved ex-post net savings for the portfolio of program claims contributed by each PA. Custom claims represent a significant portion of the entire portfolio -19 percent of electric savings claims and 56 percent of gas savings claims for PY2013 across all PAs. This incentive mechanism provides a form of extra motivation for the PAs to maximize achieved savings.

The ex-ante component of the incentive reward is designed to motivate the PAs to participate fully in the EAR process and exhibit both an effort to change and demonstration of change within the PA custom program environment. More specifically, the metrics used to score ex-ante results for custom EAR include timeliness/responsiveness, breadth of responses, comprehensiveness of submittals, efforts to proactively collaborate with the CPUC on EAR efforts, quality and appropriateness of documentation, demonstration of quality control and oversight, use of recent and relevant sources, demonstration of thoughtful consideration of CPUC input and feedback, appropriate use of DEER values and methods, and an ongoing effort to incorporate cumulative experience from the EAR process and other regulatory input. Draft ex-ante ESPI results for the custom EAR component of ESPI for PY2013 ranges from 49% to 75% of the maximum potential score across PAs. Two of the PA's scored relatively low and two of the PAs scored relatively high.

### **Applicability to Other Jurisdictions**

In an effort to address lower than desired GRRs and business-as-usual custom program implementation in California, the CPUC has greatly increased efforts to bring about change through both enhancements to make evaluation results more meaningful and actionable and through increased regulatory participation and collaboration featuring innovative EAR procedures. For the California program environment this has proved to be a worthwhile set of activities that work collaboratively to bring about change. ESPI incentives serve as a source of financial motivation to the PAs to accept the CPUC challenge of change and improvement.

This may not be a panacea for all jurisdictions, California faces conditions that include: large custom program components, high goals and a track record of relatively low GRR results, among others. Also, California has relatively high evaluation funding levels that can support the activities described in this paper.

Not all of these conditions are unique to California, based on our experience evaluating and assessing custom programs elsewhere, and we believe that there are many other jurisdictions that could benefit from the approaches described in this paper. Still, there are hurdles and barriers. For example, the funding and, closely correlated, rigor of custom evaluations is not uniform across jurisdictions. Lower rigor evaluations will tend to return GRR results that approach 1.0 for these types of programs, since a relatively large level of effort is needed to uncover often complex, project-specific, ex-ante estimation shortcomings and areas for improvement.<sup>12</sup> Conversely, M&V that features in-depth analysis of baselines, deep measurement, calibrated modeling and best practices in evaluation will tend to uncover more issues that require attention.

Jurisdictions with substantial levels of custom program funding should consider scoping evaluations that offer best practices and feature key approaches described in this paper. It is worthwhile to dig deeper and find the flaws, as well as solutions that can improve program performance. It is also worthwhile to pilot

<sup>&</sup>lt;sup>12</sup> Said another way, when evaluation engineers have limited budgets per project, they have less primary research to draw on, and are, consequently, more reluctant to overturn ex-ante savings estimates. This can lead to a higher percentage of "passed through" projects with GRRs of 1.0 in underfunded evaluations. Somewhat perversely, this may lead to lower GRRs for larger programs with more adequate evaluation budgets and, in some cases, better practices and performance. Similarly, smaller programs may have artificially high GRRs that stem in part from low evaluation budgets. Care must be taken in comparing evaluation results for custom programs across jurisdictions, particularly with respect to evaluation rigor and whether regulatory policy requires best practices on issues such as gross baseline determination.

ex-ante review activities in an effort to more carefully examine within-program processes and procedures, and to uncover weaknesses. Considering the dollars spent to implement programs that may not be functioning to full potential, the evaluation and regulatory feedback channels explored in this paper are likely to be worthwhile. A poorly functioning program can unnecessarily waste ratepayer and public funds. Through in-depth evaluation, ex ante feedback, program administrator commitment, and regulatory collaboration and incentives, an array of issues associated with inaccurate savings claims for complex custom programs can be identified and improved, thereby increasing the efficacy of energy efficiency funds.

It is notable that process evaluation represents another potential evaluation component to help improve custom program performance. However, many of the technical, engineering-focused issues are best addressed using an M&V approach that is similar to both the EAR and evaluation approach being applied in California. For jurisdictions with custom programs that don't suffer from similar shortcomings, an alternative use of evaluation resources should be considered.

Since the low GRR results described in this paper largely stem from a need for better ex-ante technical execution, the evaluation emphasis on EAR and custom evaluation is thought to be appropriate. Furthermore, evaluation results that provide a clear path to actionable recommendations is also an important element of the California custom program solution. Reporting of the frequency and impact of discrepancy factors, and the examination of differences in ex-post conclusions for various PPA elements sets the stage for identifying actionable issues and points directly to areas for improvement. EAR accomplishes the same thing, in a different setting, working directly with the PAs on projects under development, and leads to direct and actionable guidance and recommendations.

## References

- CPUC 2011. *Third Decision Addressing Petition for Modification of Decision 09-09-047*, Decision 11-07-030 July 14, 2011.
- CPUC 2013. Decision Adopting Efficiency Savings and Performance Incentive Mechanism, Decision 13-09-023, September.
- CPUC 2013. Energy Efficiency Policy Manual, Appendix E: Custom Project Review Process, Version 5, R.09-11-014, July.
- Ehrlich, C., Lutz, A., and Bradley, K., "Using Structured Desk Reviews to Manage Risks Associated with Commercial and Industrial Custom Impact Projects." *In Proceedings of the Associated Energy Services Professionals National Conference & Expo*, Orlando, FL. February 2015.
- Fagan, J., Bradley, K., and Lutz, A., 2011. "Strategies for Improving the Accuracy of Industrial Program Savings Estimates and Increasing Overall Program Influence." In Proceedings of the International Energy Program Evaluation Conference, Boston MA. August 2011
- Itron, Inc. 2010. 2006-2008 Evaluation Report for PG&E Fabrication, Process and Manufacturing Contract Group. Prepared for the California Public Utilities Commission, Energy Division.
- Itron, Inc. 2010. 2006-2008 Evaluation Report for the Southern California Industrial and Agricultural Contract Group. Prepared for the California Public Utilities Commission, Energy Division.
- Itron, Inc. 2014. 2010-2012 WO33 Custom Net-to-Gross Final Report. Prepared for the California Public Utilities Commission, Energy Division.

- Itron, Inc. 2015. 2013 Draft Custom Impact Evaluation -- Industrial, Agricultural, and Large Commercial. Prepared for the California Public Utilities Commission, Energy Division.
- Lutz, A. and Pagadala, P., 2014. "Discrepancy, What Discrepancy? Reasons for the Difference between Claimed and Evaluated Savings." *In Proceedings of the* American Council for an Energy-Efficient Economy (ACEEE) 2014 Summer Study on Energy Efficiency in Buildings, Asilomar, CA, August 2014..
- Maxwell, J., N. Gandhi, and F. Coito, 2011. "Establishing Baselines for Industrial Energy Efficiency Projects," *In proceedings of the American Council for an Energy Efficient Economy's Summer Study on Industry*, Niagara Falls, New York. August 2011.
- Rufo, Michael. 2015. "Ew, Gross! Cleaning Up Gross Baselines," *In proceedings of the 2015 International Energy Program Evaluation Conference*, Long Beach, CA. August 2015