

Overview of the Evaluations of California's 2004-2005 Efficiency Programs

Nick Hall, Carmen Best, Johna Roth, TecMarket Works, Oregon, WI
Peter Lai, California Public Utilities Commission, Los Angeles, CA
Pete Jacobs, Architectural Engineering, Boulder, CO
Lori Megdal, Megdal and Associates, Acton, MA

Abstract

In 2004-2005 California fielded over 200 energy efficiency programs funded through the public goods charge (PGC). These programs included statewide, local utility and third-party implementation efforts. Each program was to be evaluated according to the California Public Utility Commission's (CPUC's) Energy Efficiency Policy Manual and CPUC approved evaluation plans. A total of 100 independent evaluations will be completed for the program cycle. This paper presents the evaluation approaches used to address the CPUC's evaluation objectives, some of the issues addressed by the various studies, the activities used to address those issues, the evaluated program effects, and overarching findings and recommendations from the process and impact evaluations. In summary, the studies included in this paper demonstrate that the 75 percent of the kW savings predicted by the program implementers is being delivered to the state of California and that somewhat less than half of the programs are accurately projecting the impacts that they can achieve. The methods used to predict energy savings from energy efficiency programs needs to be reexamined to make the process more accurate. Process evaluations need to focus more attention on making programs more effective at capturing their energy impact objectives.

Introduction

From 2004 to 2005 California fielded over 200 energy efficiency programs funded through the public goods charge (PGC). The majority of portfolio was dedicated to resource acquisition programs (88 percent of the total approved budget), which employed a range of strategies to encourage adoption of energy-efficient technologies and practices. The remaining 12 percent of the implementation budget was for marketing, information and education programs. Programs were engaged in a variety of market sectors including the commercial, residential, industrial, and agricultural sectors and several of their sub-sectors. About 22 percent of the portfolio covered programs targeting multiple sectors, including cross-sector utility procurement efforts (10 percent) and the statewide marketing and outreach campaigns (1.2 percent).¹

Some evaluation efforts were consolidated so that about 100 separate evaluation reports are provided from the studies. Approximately 4 percent of the total portfolio budget was dedicated to evaluation. To date, 57 evaluations have been received and accepted by the CPUC, representing 14 percent of the kWh goals, 40 percent of therm goals and 15 percent of demand savings goals. These studies represent about 22 percent of the total 04-05 portfolio budget². This is because most of the large

¹ Portfolio data is derived from EEGA (Energy Efficiency GroupWare Public Access) [available 5-9-2007]. EEGA is the electronic filing process for program workbooks by utilities and third-party program implementers. Program workbooks include data on program goals, accomplishments and budgets for each utility funded effort within the portfolio.

² Many of the graphics in this paper use budget as a measure of portfolio effort to reflect the size of both information and resource acquisition programs and type of energy savings targeted.

statewide program evaluations have not yet been completed. The results from these studies will be presented during the conference if they are available.

Evaluations from all types of programs are still pending; however the majority of the third party program evaluations (non-utility non-governmental program) and the non-utility governmental programs have been completed, and are included in this analysis.

About a quarter of the program funds budgeted for resource acquisition programs have been evaluated, and about one-third of the funds for information programs have been evaluated. Most of these programs and several of the resource acquisition programs also had educational components to them, however the identification as an “information” program meant that the CPUC did not require these programs to meet energy savings goals or conduct an impact evaluation. All other programs were obligated to set and meet established energy savings (resource acquisition) goals.

Evaluations that have been completed came from a wide cross section of target markets. In Figure 1 the portion of the total portfolio effort (as measured by budget) that has been evaluated is presented. Note, for example, that all of the agricultural programs have been evaluated, and 85 percent of the residential information efforts have been evaluated yet only 19 percent of residential resource acquisition efforts have been completed to date. Again this is because much of the portfolio’s resource acquisition efforts are concentrated in the statewide programs.

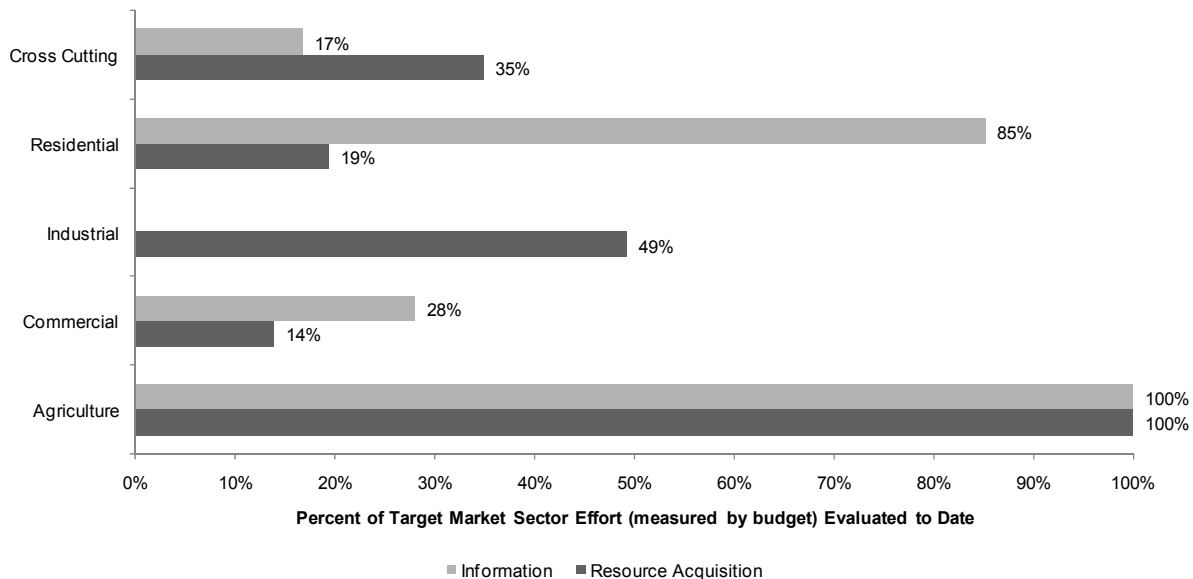


Figure 1. Portion of Target Market Effort with Completed Evaluations

**Cross Cutting includes programs that cover more than one market, utility’s procurement funds that cover multiple markets and the statewide marketing campaign.*

This paper provides a high-level summary overview of the 2004-2005 evaluation efforts as of June 2007, including the evaluation approaches used and the energy impacts of those programs.

Evaluation Approaches, Methods and Issues Addressed

The focus of the evaluation efforts varied across programs. This variation was primarily driven by the program’s focus, the implementer’s evaluation desires and the CPUC’s approved evaluation objectives. However, the available budget, in most cases, determined the rigor and comprehensiveness of the evaluation effort. The evaluation budget allowed for the 04-05 programs averaged about 4percent

of the program implementation budget and was not sufficient to evaluate all programs using reliable evaluation approaches. As a result of the lack of evaluation resources needed to reliably evaluate all programs, more rigorous (and more expensive) evaluation efforts were a priority for the larger programs. Most all information and resource acquisition program evaluations included a process evaluation component. Logically, the resource acquisition program evaluations also focused on conducting impact evaluations despite a few information program evaluations that also conducted energy impacts studies (See Figure 2).

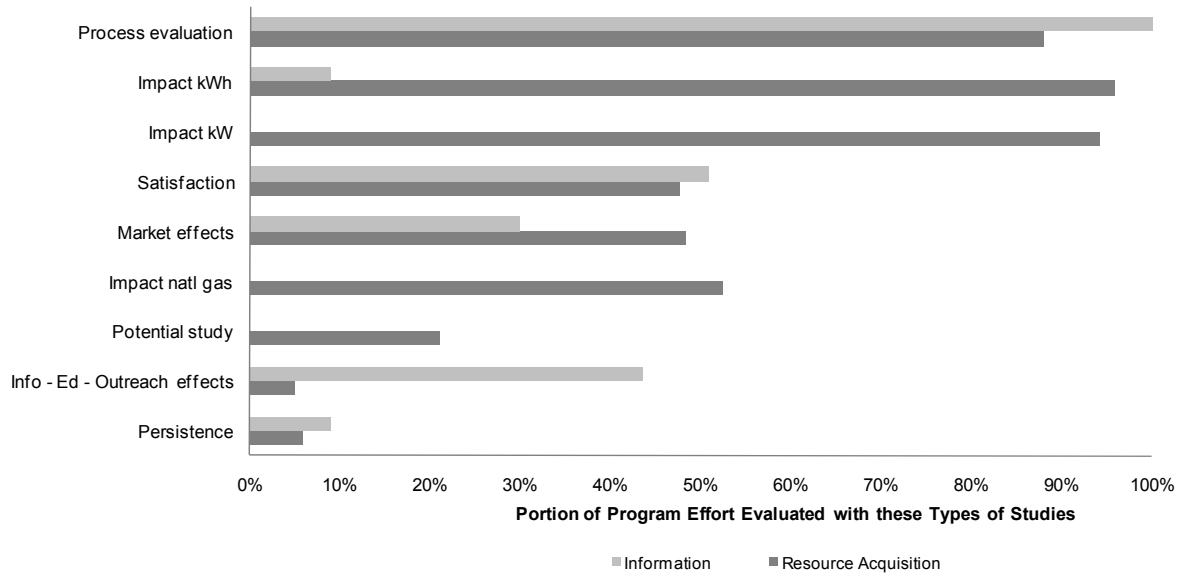


Figure 2. Types of Evaluations Conducted

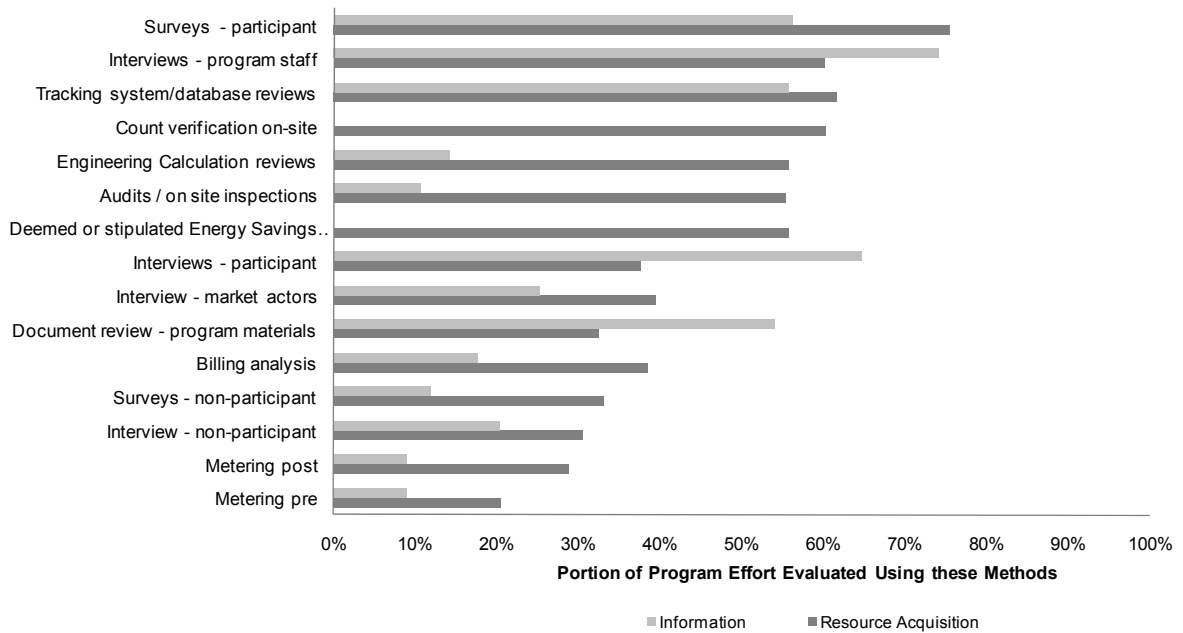


Figure 3. Evaluation Methods

**Program Effort is represented by the approved budget; Specific methods used to adjust kW, kWh and therm, impact estimates are provided in a later section of this paper.*

As expected, the evaluation methods employed are closely related to the type of evaluation that was conducted. Figure 3 illustrates that while both resource acquisition and information program evaluations relied heavily on participant surveys, resource acquisition program evaluations also were more apt to use on-site inspections and audits to evaluate energy savings. However, while the CPUC Policy Manual required the use of IPMVP evaluation approaches (which require on-site instrumentation applied measurements), many evaluations did not conduct on-site or site-specific verification, monitoring or metering activities. In fact only about 60 percent of the impact evaluation efforts, in some way, complied with this requirement. The primary reason for not complying with the CPUC-ordered IPMVP approach was that the programs did not budget enough resources to conduct the field efforts associated with the on-site measurement requirements embedded in IPMVP. Few of the third party program administrators understood the requirements associated with IPMVP efforts, and did not plan evaluation resources to accomplish the IPVMP objectives. After the CPUC approved the program implementation budgets, many program administrators were reluctant to move funds from implementation efforts in order to enable the IPMVP evaluation requirement. Following this experience the CPUC increased the evaluation budgets in the next program cycle to 8 percent in order to obtain more reliable studies.

Measuring Energy Impacts

Based on program implementation plan savings targets from the suite of California’s 2004-2005 programs, the portfolio was anticipated to have annual savings of 3,746,785 MWh, 55,724,451 therms and demand savings of 760 MW. Lifecycle saving goals added up to 43,531,173 MWh and 695,549,999 therms.³ Of these originally anticipated savings, 14 percent of the annual kWh and kW goal, and 39 percent of the therm goals have been evaluated to some degree in the 36 impact evaluation reports completed and reviewed as of this paper. Just over half of the evaluation reports have been received, and only 1 of 9 statewide impact evaluations have been received. Only savings that have been evaluated and approved by the CPUC are included in this paper.

Completed impact evaluations report actual annual program savings of 79.02 MW of demand savings, 345,745 MWh and 7,479,409 therms.⁴ The annual evaluated savings were compared to the annual program goals and reported energy savings which were based on data available from the CPUC’s program progress and reporting tracking system. The following table presents the results as the percent difference between the program-reported savings and the evaluation confirmed savings. This table indicates that while the programs reported savings very close to their goals, the savings confirmed via the evaluation efforts are somewhat less than expected by the programs. This table indicates that of the programs for which evaluation studies have been completed implementers report that they have achieved 91 percent of their kW objectives, however according to the evaluation reports, about 75 percent of the kW objectives have been achieved.

Table 1. Program Reported Savings v. Evaluation Reported Savings

	kWh Annual	Therm Annual	kW
Percent of Annual Goal Achieved based on Program Reported Savings	94%	90%	91%
Percent of Annual Goal Achieved based on Evaluation Reported Savings	68%	41%	75%

³ EEGA (5-9-2007)

⁴ Definitions used for “peak demand” were not consistent across evaluations.

Figure 4 graphically represents the compared savings by program type. The light gray columns represent the percent of the *program-reported* annual savings compared to the program annual goal as reported in the reporting system, while the darker columns provide the *evaluation-reported* annual savings as a percent of the annual program goal. The difference between the energy savings reported by the programs, and the energy and peak demand savings documented by the evaluations is significant. The statewide difference represents only one study and is not representative of the other statewide programs. In general, this graphic indicates that overall the programs are providing reasonable estimates of kW impact with the third party programs projecting demand impacts the most accurately. The kW impacts projected by the local IOU programs indicate that they have met 81percent of their projected kW goals, however the evaluation efforts report that about 67 percent of the kW objectives have been acquired. The programs operated by local governmental units report that they have obtained 94 percent of their kW goals, however the evaluation efforts indicate that they have acquired about 87 percent of their objectives.

Demand reduction is an important component of energy efficiency programs, and the majority of impact evaluations in this cycle did include estimates of demand reduction. However, definitions of “peak demand” were not consistent across the evaluations. The CPUC policy manual defined peak demand as Noon to 7 p.m. Monday through Friday, June, July, August, and September. Only three evaluations included that specific definition in their report of demand savings. Three others included a definition that is close, and the remainder either did not provide a definition or included an alternate definition that allow all kW reporting to be comparable. For the up-coming 2006-2008 evaluation cycle the definition has been grounded in required protocols to make sure that the kW definition is well defined along with the specific dates over which the definition applies. In addition, the new protocols require specific reporting formats.

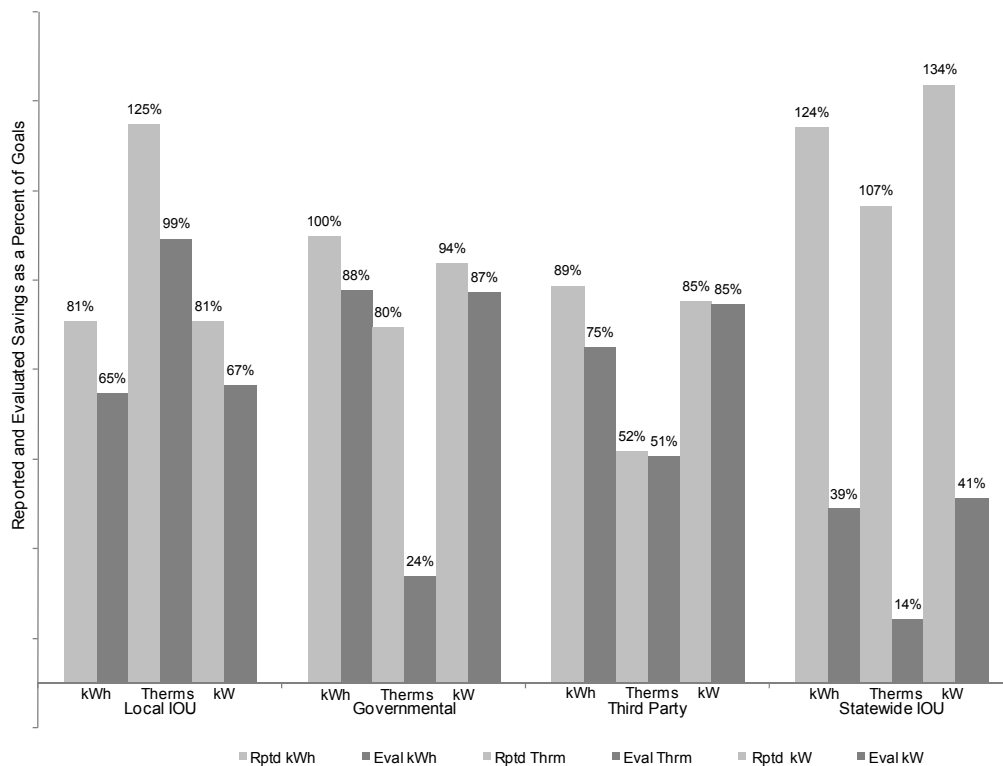


Figure 4. Reported and Evaluated Annual Energy Savings versus Goals (by Program Type*)

*The one statewide evaluation was of the Multifamily Energy Efficiency Rebates program. This program had a low realization rate primarily due to adjustments to deemed lighting saving estimates, programmable thermostats and adjustments to boiler control savings.

Because of an effort by the CPUC, estimates of lifetime energy savings were reported by all evaluators that submitted the required energy impact tables. This was a significant improvement in the availability of this information from the evaluations in the 2004-2005 evaluations compared to the 02-03 evaluations. Accurate lifecycle savings estimates are essential for designing effective programs and for developing public policy regarding energy efficiency resources. For the evaluations completed to date, the programs' ability to meet lifecycle goals was equivalent to the achievements of annual energy savings goals as presented in

Figure 4.

The standard impact evaluation techniques used to adjust the estimated savings is presented in Figure 5. The methods used are presented as a portion of the evaluated kWh, therm and kW energy savings goals to reflect relative program size. Note that review of nearly all of the kWh, therm and kW savings evaluated to date have had, at a minimum, an installation verification process. The majority used on-site verification alone or coupled with phone verification activities. This is a significant improvement over the 2002-2003 evaluations, but not yet rigorous enough to have confidence in all of the study findings. Multiple combinations of additional field information including engineering algorithm review, metering, utility consumption data, and survey information were also used to evaluate and adjust both energy (kWh 93 percent; therms 92 percent) and demand savings (87 percent). Engineering algorithm review was the most common adjustment and was most commonly based on field and on-site data (60 percent), followed by a combination of secondary and field information (33 percent). Only 8 percent of savings evaluated with an engineering algorithm review relied on a "desk/paper reviews" without on-site examinations to confirm at least some part of the information being reviewed. In the savings evaluated via the use of consumption data, the data was most typically used in regression analysis savings assessment approaches.

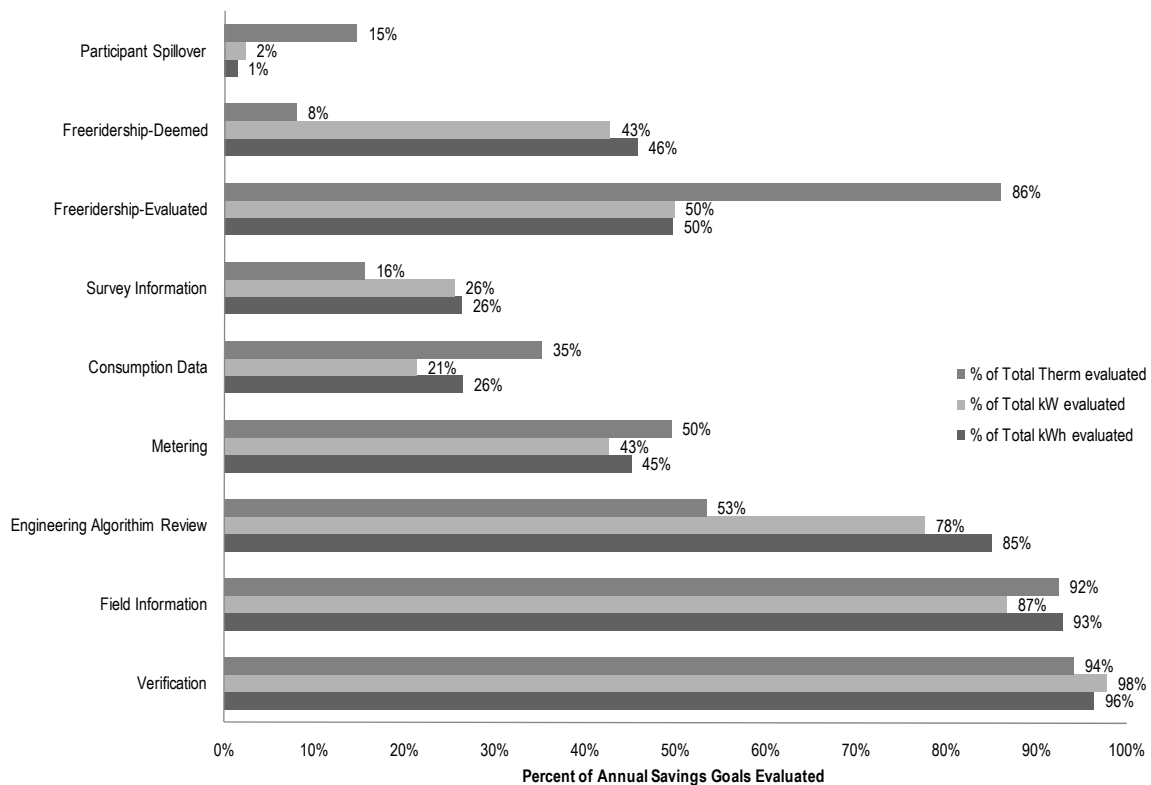


Figure 5. Evaluation Adjustments to Reported Energy Savings

Half of the kWh and demand savings included in the completed impact evaluations included an evaluated freeridership adjustment, while 85 percent of therm savings included an evaluated free ridership adjustment. For the remaining portion of savings (kWh, 46 percent and kW 43 percent) evaluators applied the CPUC's deemed NTG adjustments. This is a significant weakness of the studies because the deemed NTG adjustment may not apply to the type of program implemented or the program's delivery approach, despite evaluator's attempts to validate the use of the deemed NTG adjustment. In many cases there were insufficient evaluation funds to conduct a full NTG analysis. Evaluated freeridership adjustments were all made using survey data gathered from participants using questions that followed the general types of approaches that have become standard in the industry. These questions basically focus on asking participants (and in some cases non-participant) questions about their equipment and participation decision process and how the program has influenced that process. However, the estimates of freerider levels has become a subject of evaluation discussion about the ability to accurately measure freeridership, especially in California where there has been extensive focus on energy efficiency for an extended period of time. It is now unclear if participants are able to identify their program-influenced decisions from their non-program influenced decisions.

Eleven evaluation reports provided confidence intervals around the overall saving estimates. This represents 30 percent of kWh, 35 percent of kW, and 55 percent of the therm energy savings goals evaluated to date. This again was a significant improvement over the 2002-2003 evaluation, but not yet occurring at an acceptable level. Nearly all of these estimates provided were at the 90 percent confidence level. The evaluation protocols for 2006-2008 require this analysis and reporting for all impact evaluations in which the intervals can be calculated (supported by the study approach).

Process Evaluation

While the session in which this paper is presented focuses on kW impact impacts, we also summarize the process evaluation efforts completed to date to allow for a more rounded presentation of the total evaluation efforts. The topics addressed in the process evaluations varied to a certain extent in accordance with the type of program offered, however, many of these issues could be addressed within the process evaluation efforts for any program. In 2004-2005, program evaluations focused on program operations and implementation systems, effectiveness of program theory and design, and other program elements that had an effect on the success of the program. Figure 6 presents the process issues examined in the evaluations and indicate if the evaluation made a recommendation to improve the program. While the number of programs evaluating process issues is of interest and importance, the comprehensiveness and resulting recommendations resulting from thorough evaluations of these issues is of even greater significance. One of the key objectives of the process evaluation is to make recommendations to improve the programs being evaluated; this is especially true for resource acquisition programs that need to acquire kW impacts.

Across all of the process evaluation issues examined, evaluators made recommendations for improvements in about half of the studies. This was somewhat disappointing in view that the production of these recommendations is the primary purpose for conducting a process evaluation. However, in those cases where no recommendations were made, comments from the evaluators indicate that the program was generally successful in the areas investigated. One of the underlying problems with many of the third party process evaluation was the level of in-depth investigation conducted. Third party implementers tended to budget their process evaluations in a way that only allowed for a limited investigation into the operations of their programs. The purpose of a process evaluation however, is to

conduct rigorous in-depth assessments of one or more components associated with the program in order to make recommendations that improve the operations and cost effectiveness of that program. However, because of the evaluation budget restrictions, the process evaluations sometimes had to rely on more general investigative efforts that were not capable of serving as the foundation for recommended program operational changes. Figure 6 presents the areas of investigation associated with the 04-05 process evaluation efforts, and identifies if the study also recommended changes associated with that area of investigation. The X axis of the graphic presents the percent of the process studies that focused on each area of investigation included on the Y axis. The color of the bar indicates the percent of the studies that provided a recommendation (dark bar) within each area of focus that is expected to improve the program's operations relative to that area of investigation.

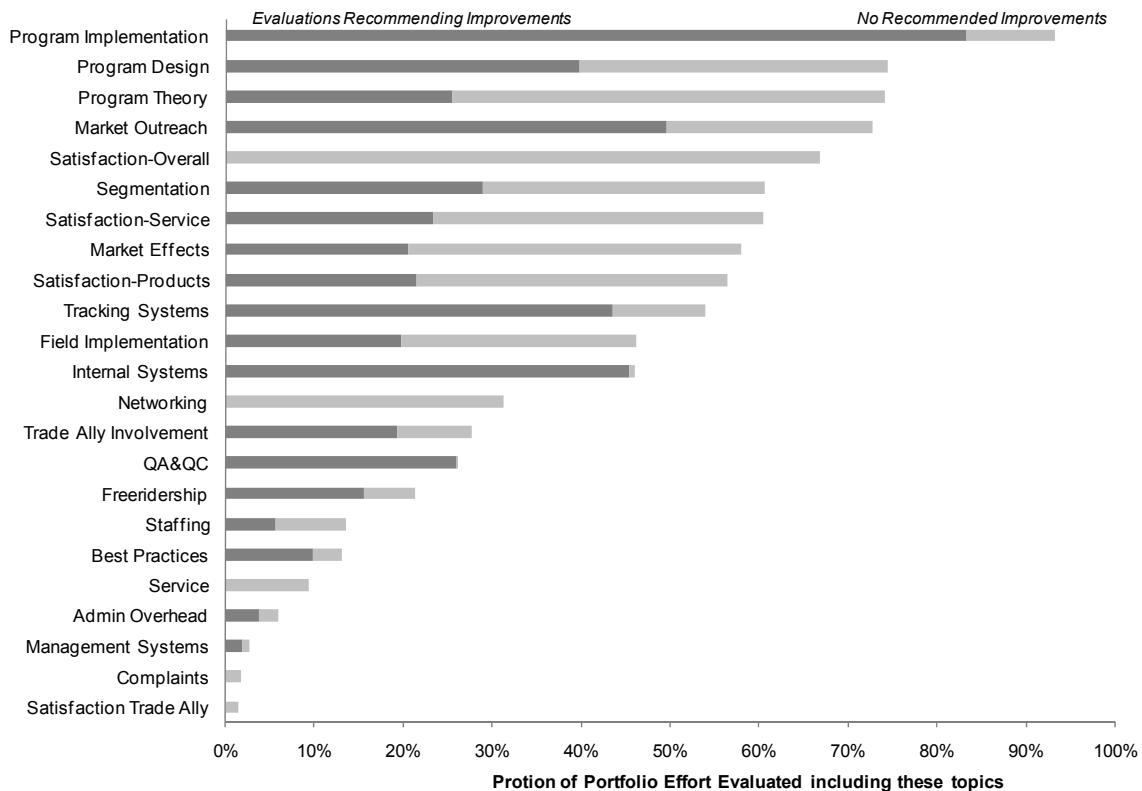


Figure 6. Process Evaluation Recommending Improvements

Overall Findings and Recommendations

As one of the most valuable components of the program evaluations, the findings and recommendations provided by evaluators provide insight to decision makers on whether or not a program should be continued, the accomplishments and recommendations for improving these programs to achieve energy saving or other goals.

Since these evaluations were conducted at the program-level, the impact and process recommendations focused on programs rather than types or groups of programs. High-level findings and recommendations were aggregated to determine if there were common themes that emerged from the program evaluations. Evaluator's recommendations on the continuing need for the program was required in the Policy Manual Objectives and as a result the majority of evaluators provided a specific recommendation to continue or not continue a program. As noted in the following table, two resource

acquisition and four information only programs were not recommended for continuation for a variety of reasons including their inability to achieve their energy impact goals, the validity of their program theory and the opportunities available to overcome market barriers or achieve market penetration.

Table 2. Recommendations for Program Continuation (57 evaluations completed)

Type of Program	Continue	Discontinue	No recommendation**	Evaluations Complete	Program Budget Evaluated
Resource	85%	3%	12%	36	\$160 M
Information	82%	10%	8%*	21	\$36 M

*One information only evaluation recommended further research before continuing or discontinuing the program.

**No summary recommendations were offered.

The programs not recommended for continuation represented a mixture of IOU and third party programs. Three programs were not recommended for continuation because of flawed theories or market assumptions. In these studies the evaluation contractor concluded that the market assumptions made by the program, for which their programs were designed to change, were not valid. That is, the market did not operate in the way the program designers assumed, rendering their program theory invalid. Another program was recognized as a highly effective program that had reached its market saturation point of eligible customers and needed to expand program eligibility criteria and change their delivery offerings / approach or face diminishing returns. Two programs were considered acceptable, however the evaluation contractors suggested a market-based delivery approach would be more effective than a program participant based approach. For these programs the evaluation contractors recommended significant program design and delivery changes that are expected to make the programs viable, but did not recommend continuation of the program in their current design and delivery approach. Another of the information and education evaluation reports recommended additional research needed to be conducted before a recommendation to continue or discontinue the program could be reached.

The evaluation contractors provided key process and impact evaluation findings across a wide range of topic areas as indicated in Figure 7. A large share of the key findings on program approach were positive, citing the program's general effectiveness toward accomplishing their objectives or in their operations. This occurred across both information and resource acquisition programs and indicates that most programs are on the right course and are achieving significant savings. The most commonly cited problem (6 mentions) was with program management or other significant implementation challenges. With respect to impact goals, 19 out of the 36 (53 percent) resource program evaluations noted in their findings that the program did not meet its energy savings goals. This is a very significant and troubling finding particularly for the third party and government partnership programs which make up the largest portion of this sample of evaluations. Because about half of these programs are not reaching their energy impact objectives, these findings suggest that the way in which the programs are projecting their accomplishments needs significant attention. It also suggests that the cost effectiveness assessment process used to determine what programs should be funded needs to be revisited. Far too many programs that are projected to be cost effective are not actually delivering cost effective savings. In most cases the evaluation resulted in changes in the amount of energy savings used to project cost effectiveness; however, there were other programs in which the implementation efforts or the duration of the program was a barrier to achieving the projected energy impacts. Yet these findings also demonstrate that a good number of programs, but less than half, are accurately projecting the level of savings that they can be expected to achieve. The good news is that some programs are reliably projecting their energy impacts, however, there is significant room for improvement in the ability to accurately project energy impacts.

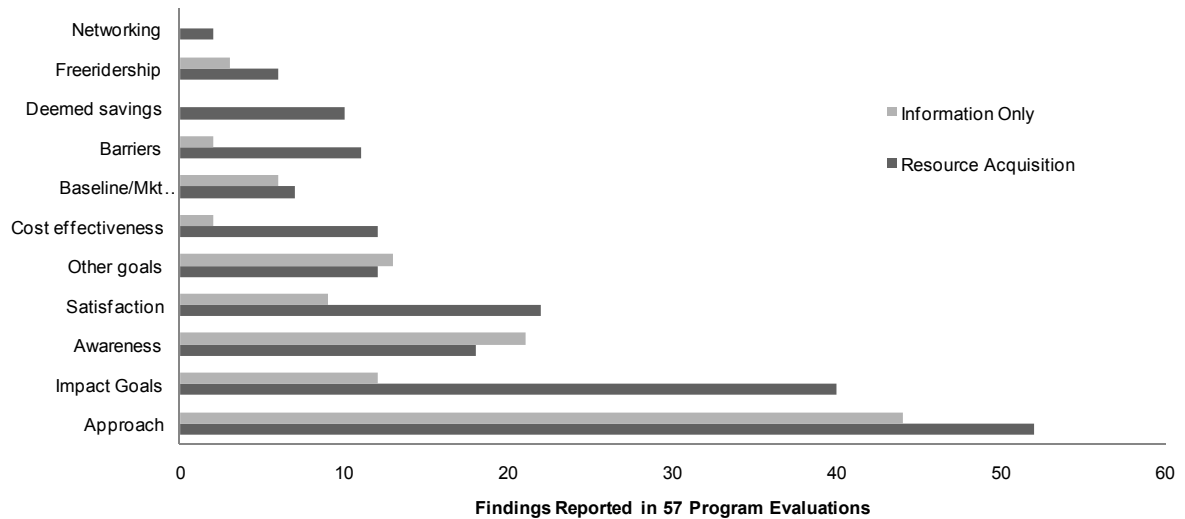


Figure 7. Key Findings

Improvements in public awareness due to the program were cited for many of these programs (25 mentions). This suggests that many of the program’s outreach and marketing efforts are effectively reaching customers, but there is still room for improvement. Several process evaluations also cited problems and barriers to participation due to a lack of understanding about the program and its benefits, and confusion in the market about what programs and services are available and what is needed to participate. This finding was not limited to the program-specific evaluations, but was also evident in the statewide marketing and outreach evaluation that found the messaging provided was not significantly actionable, and acted in many cases to inform citizens about issues on which they were already aware. Clearly the programs can benefit from more effective marketing and outreach. However, studies also noted that persistence of measures was an issue impacting energy savings, suggesting that programs need to better understand the environments in which measures are placed and the use conditions of those measures. This is particularly true for retrocommissioning programs in which savings can erode as fast as the next need to over-ride building management software settings.

Recommendations for improvements were typically program specific; however, there were some common areas of recommended change across several studies. These are presented in Table 3.

One of the most frequent recommendations is to adjust the deemed savings estimates used by the program to better reflect the program, market and use conditions. This is consistent with the finding that many programs were not able to achieve their energy savings goals as a result of inaccurate savings estimates. Several recommendations also highlighted the need to coordinate their services or marketing efforts with utilities and local resources.

Table 3. Summary of Common Recommendations (numbers of recommendations by category)

Category	Recommendation	Resource Acquisition	Information
Approach and Awareness	Coordinate with local resources	2	7
	Simplify the program application and process	4	3
	Improve materials/resources	4	2
	Provide actionable recommendations	1	5
	Obtain more staff	5	1
	Improve training	2	3
	Widen base of intervention	2	2
	Get more information to customers	2	2
	Coordinate with utilities and existing programs	4	-
	Expand eligibility criteria	3	-
Impact Goals	Adjust/validate deemed savings	17	2
	Conduct additional research to determine impacts	1	2
Tracking	Improve tracking methods	4	6
	Document participants in the program	-	2
	Encourage proper program tracking	-	2
Outreach	Improve materials/resources used	2	2
	Widen base for program outreach	2	1
Technologies	Review and adjust available technologies	6	-
Program Length	Program needs more time to operate	3	2
	CPUC should eliminate contracting delays	1	3

Conclusion

The 2004-2005 Public Goods Charge program evaluation effort in the state of California was much improved over that of 2002-2003. This was due to several factors. First the CPUC and the Master Evaluation Contractor Team were able to provide more oversight than permitted under the 2002-2003 evaluation efforts, particularly to studies associated with the third party programs. Third party implementers also had a better understanding of the cost of the evaluation efforts needed as compared to their 2002-2003 budgets. Finally the results from the 2002-2003 evaluations demonstrated that under-budgeted evaluation does not provide as reliable results because it cannot employ the rigorous evaluation approaches necessary. The consistency, rigor and value of the 2004-2005 reports are certainly a step in the right direction. The findings of these evaluations are an important step in improving the implementation of future programs, as well as the culling of ineffectual, non-cost-effective programs from the portfolio.

The primary message that emerged from an aggregated look at the program evaluations is the need for more accurate stipulated energy savings to develop realistic program goals that supports more accurate estimations of cost effectiveness. An additional high-level outcome from the evaluation process is that cost-effective programs do not just happen, there must be strong, vigorous and constant program management and oversight to maintain cost effective programs. California is planning and depending on energy efficiency to meet a significant portion of its future energy needs as well as

achieve reduce greenhouse gas emissions. Accuracy in the planning, implementation, and final evaluation is critical to achieving these goals.

The 2006-2008 program cycle is underway and the evaluation efforts for this cycle, which is arguably the largest energy efficiency portfolio in the world, will be more precise and more comprehensive because it will be guided by the California-specific Evaluation Protocols and the associated California Evaluation Framework. This paper presents a brief summary of the evaluation results from the 2004-2005 studies. A more complete presentation and discussion of these findings can be found the Master Evaluation Contractor Teams Cross Cutting Meta Evaluation to be completed in November of 2007.