Final Report

California 2002-2003 Portfolio Energy Efficiency Program Effects and Evaluation Summary Report

January 16, 2006

Prepared for

Southern California Edison
and the Project Advisory Group

Prepared by
Nick Hall and Carmen Best, TecMarket Works
Pete Jacobs, Architectural Energy Corporation
Lori Megdal, Megdal & Associates

TecMarket Works
165 West Netherwood Road, Suite A, 2nd Floor
Oregon, WI 53575
Voice: (608) 835-8855

Table of Contents

Table of Figures

Table of Tables

Executive Summary

This report presents, discusses and summarizes the efforts and findings of the evaluations of the California 2002-2003 energy efficiency program portfolio (2002-2003 Portfolio) funded by the California Public Goods Charge (PGC). The eighty-five programs offered in the 2002-2003 Portfolio were implemented by three different types of entities at two different levels. The investor-owned utilities (IOU) were responsible for programs at the both the statewide (SW-IOU) and the local (Local-IOU) levels, while only local programs were implemented by local governments (Local-GOV) and non-utility non-government third party implementers (Local-TPI).

This report, while planned to be completed in 2004, took considerably longer to develop than expected. Several of the 2002-2003 Portfolio programs were provided with lengthy extensions, delaying the completion of their - and the present - evaluation efforts. Several other evaluation reports were delayed for a variety of reasons. As of the writing of this report, five evaluations have yet to be finalized and made available for this assessment, not including the 2003 statewide and local IOU program evaluations. The programs assessed in these missing evaluations represent about 2 percent of the total 2002-2003 Portfolio expenditures and less than 1 percent of the 2002-2003 Portfolio kWh energy savings goals, but represent about 20 percent of the natural gas savings goals. Two of these programs have been extended and, as a result, the evaluations have not yet been completed. For other reasons the evaluations of the remaining three programs have not been completed as of the publication date of this report. Where available and applicable, this report includes information from these five programs based on their planned evaluation efforts and information presented in their final program reports. The remaining and majority of the material included in this report is taken directly from the program evaluation reports completed, the program implementation plans provided to the California Public Utilities Commission (CPUC) and the final program progress reports provided by the program administrators.

At the time of this report, not all of the evaluations of the 2003 statewide and local IOU programs were complete. For consistency in analysis, none of evaluation findings for the 2003 IOU programs are included in this report. However, the 2003 statewide and local IOU programs are similar to those run by the IOUs in 2002 which did undergo evaluations. (For program year (PY) 2002, local and statewide IOU programs were funded for one year, while the non-utility programs were funded for a two-year period (PY 2002-2003)). In order to estimate the energy impacts for the 2003 statewide and local IOU programs reported in this document, we applied the program-specific realization rates from the 2002 evaluations to the reported 2003 ex-ante savings in order to include two years of effects information for the statewide and local IOU programs. The third party and local government program evaluations include effects from both 2002 and 2003. This effort allows the comparison of effects across IOU and non-utility programs and provides for a more comprehensive report.

Two basic program classifications are used in the present study, those that are designed to directly acquire energy resources (resource acquisition programs), and those officially classified as “information-only” programs. Information-only programs are generally designed to help customers make energy-efficient product purchases, inform customers about energy-efficient

options available in their home or business, or help channel customers into one or more of the resource acquisition programs. Most information-only and several of the resource acquisition programs also have educational components, however, the term "information-only" indicates that the CPUC did not require these programs to meet energy savings goals. All other programs are expected to set and meet established energy savings, or resource acquisition, goals.

The 2002-2003 Portfolio included fourteen statewide IOU programs (eight resource acquisition and six information-only, including the Codes and Standards program) and sixteen local IOU programs (four resource acquisition and twelve information-only). Local governments implemented seven local programs in the 2002-2003 Portfolio, three of which were resource acquisition programs and four information-only. The majority of 2002-2003 Portfolio programs were implemented by non-utility, non-governmental third parties. Of these forty-eight local TPI programs, thirty-nine were resource acquisition programs and nine were information-only. (Table 1)

Table 1 Summary of 2002-2003 Programs

	Resource Acquisition	Information-Only	Total
SW-IOU	8	6	14
Local-IOU	4	12	16
Local-GOV	3	4	7
Local-TPI	39	9	48
Total	54	31	85

The 2002-2003 Portfolio programs covered a wide range of markets and market actors. Several programs, including many of the information-only programs, were crosscutting, addressing multiple markets.

Summary of Findings

The evaluations of the 2002-2003 Portfolio programs, and consequently the overall evaluation of the 2002-2003 Portfolio, generally suffer from three key problems: incomparability, incompleteness and a lack of evaluation rigor. Despite the overall weakness of the evaluation effort at the portfolio level, there are several program evaluations that employed rigorous methodological approaches, providing reliable savings estimates. However, these studies are offset by others that employed weak methodologies, some of which the TecMarket Works master evaluation contractor team (Master Evaluation Team) considered to be more planning estimation approaches than actual evaluations. As consolation, the most rigorous studies are generally those for the programs with the largest savings, which means that a high percentage of total claimed savings were accurately and adequately evaluated for the 2002-2003 Portfolio, and represent a relatively accurate picture of the actual 2002-2003 Portfolio's savings.

The total evaluated energy savings for the 2002-2003 Portfolio are summarized in Table 2.

Table 2. Summary of Evaluated Energy Savings for 2002-2003 Portfolio

	Program Type	Number of Programs Evaluated	Evaluated kWh Savings	Evaluated Therm Savings
Resource Acquisition	Local-GOV	3	4,280,328	0
	Local-IOU*	4	26,632,698	3,215,710
	Local-TPI	35	172,149,388	4,295,646
	SW-IOU*	8	1,525,779,743	23,823,385
	Total	**50**	**1,728,842,157**	**31,334,742**
Information-Only	Local-GOV	4	0	0
	Local-IOU*	11	2,398,807	0
	Local-TPI	9	0	0
	SW-IOU*	6	243,941,374	5,736,632
	Total	**30**	**246,340,181**	**5,736,632**

* In order to estimate the energy impacts for the 2003 statewide and local IOU programs reported in this document, we applied the program-specific realization rates from the 2002 evaluations to the reported 2003 ex-ante savings in order to include two years of effects information for the statewide and local IOU programs.

Throughout this report discussion of evaluation efforts, in which energy savings are the mode of comparison, the unevaluated program goals[1] are used as the baseline not the evaluated energy savings. Energy savings goals were used as the baseline to provide a more balanced comparison between programs, since the rigor of evaluations was highly variable across the portfolio.

The use of the 2002-2003 Portfolio evaluation reports as a resource planning tool, however, is seriously hindered by the above-listed problems across the 2002-2003 Portfolio evaluations. It should be noted however, that these studies were conducted during a period in which evaluation efforts were not guided by an officially adopted protocol, but rather by the Energy Efficiency Policy Manual (EEPM)[2], which provides limited guidance in this area. These studies were also conducted during a period of time in which the typical evaluation budget was between 4 and 5 percent of the program budget and there were a significant number of small programs (for which the fixed component of evaluation costs made 5 percent inadequate to conduct an appropriately rigorous study).

Incomparability

The incomparability of the evaluations, and therefore of the evaluation-confirmed program achievements, makes it difficult to draw reliable conclusions about the contributions of the 2002-2003 Portfolio to California's energy needs. Definitions, evaluation requirements, necessary rigor and general reporting should be consistent across the evaluation efforts to ensure result comparability. The incomparability problem will likely improve in the 2004-2005 program evaluations given the recent efforts to clarify evaluation guidelines, and the addition of more oversight and direction in the evaluation planning process.

For example, the CPUC's internal evaluation, measurement and verification (EM&V) plan review and approval process document for 2004-2005 resource acquisition programs includes

[1] Program goals are also referred to as "anticipated energy savings"

[2] *Energy Efficiency Policy Manual*. Prepared by the Energy Division, California Public Utilities Commission. November 29, 2001.

language to help ensure that evaluation will provide more thorough energy and demand savings estimates. More specifically, it states:

> To meet the CPUC evaluation requirements, the EM&V Plan must have the evaluation:
>
> a. Provide peak savings.
>
> b. Provide first year kWh and kW savings along with annual savings for years 1 through x depending upon an assessment of measure lifetimes for this program.
>
> c. Measure net-to-gross and provide net energy and peak savings.

The adoption of evaluation protocols that prescribe both evaluation rigor and reporting requirements should serve to essentially eliminate the problem of incomparability for 2006 and future programs.

Incompleteness

The 2002-2003 Portfolio was not completely comparable because several evaluations were incomplete in their reporting of key metrics.

While most of the program evaluation studies report first-year savings, the majority exclude an assessment of the expected lifetime of those savings or an estimate of the total program savings over time. Some evaluations made comments on effective useful life (EUL) estimates and modified these to feed revised cost-effectiveness calculations, but most did not provide projections on the lifetime impact of program efforts. The exclusion of lifetime savings analysis from the evaluations seriously hampers the usefulness of the studies and significantly restricts their ability to inform public policy or energy supply decisions. Evaluation-confirmed annual savings over the lifetime of impacts is needed to establish an as-delivered portfolio total resource cost (TRC) from independent evaluation efforts.

Program-level cost-effectiveness is also evaluated inconsistently across the 2002-2003 Portfolio. Only 18 of the 50 (36 percent, and representing only 8 percent of 2002-2003 Portfolio expenditures) resource acquisition program evaluations reported evaluation-based cost-effectiveness (TRC). This, along with energy and demand savings, is essential information for assessing re-investment potential and conducting portfolio planning activities. Prior to California's adoption of the 1998 EM&V Protocols, many impact evaluations did not include cost/benefit-analyses, but the IOUs did use the evaluation results to conduct ex-post cost/benefit assessments. Evaluation-based cost-effectiveness tests should be provided within the independent evaluation report so it is readily available to support overall portfolio assessment and comparisons to the projected cost-effectiveness provided by the program implementers and administrators.

To be fair, the EEPM in effect at the time of the 2002-2003 program evaluations could be viewed as ambiguous in the area of impact metric definitions. Its Section 6 on EM&V requires "[m]easuring the level of energy and peak demand savings achieved,"[3] but does not specify whether these are first-year or lifecycle savings. It also calls for an analysis of cost-effectiveness which requires an estimate of lifecycle savings. Yet, the EEPM also provides assumed lifecycle

[3] Energy Division, California Public Utilities Commission, 31.

and net-to-gross ratios to be used for the program proposal filing. The EM&V section does not explicitly state that these factors need to be measured and used in the evaluations. The final project workbooks all calculate lifecycle savings, which may have been interpreted as satisfying the EEPM requirement. If an evaluation plan omitting cost-effectiveness analysis was approved by the CPUC, it could reasonably be interpreted that performing cost-effectiveness calculations was not required of the evaluation contractor. Year-by-year reported savings were not required by the EEPM or calculated in the program workbooks, so it is logical that none of the evaluations reported results in this manner.

Lack of Rigor

A surprising lack of rigor was found in the evaluation reports as a whole. A full 29 percent of the resource acquisition 2002-2003 Portfolio did not have their energy savings adjusted from estimated gross savings. While the remaining evaluations employed a wide range of evaluation approaches, some studies reported ex-post impacts by applying per-measure deemed or program-estimated savings to the program's tracking system measure counts. This methodology should not be considered an evaluation effort, but instead an accounting approach for confirming estimated savings projections. Other studies used more rigorous approaches that can be expected to provide more accurate and reliable estimates of energy impacts.

The EEPM prescribes a fairly rigorous evaluation standard in its required use of IPMVP (International Performance Measurement and Verification Protocol) evaluation approaches (which require on-site technology-based measurements). However, many evaluators ignored this requirement and did not conduct on-site or site-specific verification, monitoring or metering activities, citing lack of evaluation funds, even though administrators were instructed to budget for IPMVP-based evaluations. In fact only a slight majority of the kWh energy savings (55 percent) was evaluated with sufficient rigor to provide impact estimates with a basic level of reliability. Only 23 percent of natural gas and 44 percent of demand savings were evaluated using reliable methods.

Accurate impact evaluations require, at a minimum, either the implementation of the IPMVP requirement of establishing savings estimates with field-measured components or a consumption analysis that includes pre- and post-program utility bills from participants. Few of the 2002-2003 program impact evaluations met this rigor requirement.

Conducting net-to-gross adjustments is also important for projecting accurate saving estimates. Across the portfolio of resource acquisition programs, only twenty-one program evaluations (42 percent) included an analysis of free-ridership instead of accepting the default net-to-gross ratio. Less than half (23 of 50) of the program evaluations took free-ridership into consideration when reporting savings, covering only 28 percent of kWh, 29 percent of demand and 45 percent of natural gas savings goals.

Very few of the evaluations examined any of the different types of net-to-gross adjustments that typically act to increase savings estimates. For example, only one of the reports included the impacts associated with the program changing the way the market operates. Other positive net-to-gross components occasionally measured were participant spillover (2 reports), non-participant spillover (2 reports) and education and information influence (1 report). Where these

positive adjustments are not included in the study and free-ridership estimates are, savings estimates may be substantially conservative if the programs had any measurable effects on the operations of the markets in which they are placed - a likely result for larger programs and programs involving education, advertising or training.

Many of the evaluation methodologies, especially for the smaller programs, were hampered by evaluation budgets that were too low to support rigorous or reliable estimates of net energy impacts.

If the post-2005 draft evaluation protocols (currently in the public review process) are approved, future evaluation summary studies can report the confidence levels around evaluation-reported program saving estimates and provide a comparative discussion of the potential sources of bias and an assessment of the reliability of the estimates. These metrics were not required and so are not reported in the evaluations of the 2002-2003 Portfolio programs. This will most likely hold true for the 2004-2005 program studies as requirements and budget levels remain similar to those for the 2002-2003 evaluations.

The original RFP and Work Plan for the present project included a significant effort for a summary analysis, comparative analysis and a possible meta-analysis. Yet, the evaluations were too often not comparable and not rigorous enough to support these types of analyses.

Lessons Learned and Recommendations

A number of valuable lessons were learned during the course of the present project. The majority can be classified as either related to evaluation implementation and reporting protocols or more administrative, including areas like contracting, budgeting, and planning.

A summary of these and related recommendations include the following:

1. **Develop prescriptive evaluation protocols to guide the evaluation budgeting and implementation processes**. The EEPM was not designed to guide evaluation efforts and tends to be ambiguous in this area. Evaluation contractors and implementers, particularly the TPI contractors, clearly would have benefited from help in understanding the evaluation requirements of the EEPM. This was especially true for complying with IPMVP field procedures, including that all IPMVP options require some form of technology-based on-site field measurements (field surveys and interviews do not comply with IPMVP requirements). Additionally, the budgeting process would have benefited from implementers having a clearer understanding of the complexity of the evaluation process and what resources are required to provide reliable evaluation results. While providing assistance in these areas is one option, a better approach is to abandon the use of the EEPM as an evaluation guidance document and adopt protocols specifically designed for guiding evaluation efforts.
2. **Establish clear definitions of the evaluation metrics required from the evaluation process, including those for energy savings (kWh and therms), demand impacts (kW), lifecycle or lifetime savings and evaluation-confirmed net-adjusted savings**.
3. **Establish clear required evaluation protocols, especially for reporting and estimating net ex-post energy impacts and for conducting process evaluations**.

4. **Establish impact estimation protocols and sampling requirements.** The metrics and protocols called for in 2, 3 and 4 above are critical if energy efficiency programs are to be relied upon to help plan for California's energy needs. Currently, evaluations vary widely in their levels of rigor, comparability and completeness. Policy makers are significantly hampered when evaluations exclude assessments of ex-post kW effects, limit the reporting to only the first-year savings or exclude estimates of the rate of savings degradation (kW, kWh, and therms). Including program-specific free-rider estimates and participant spillover would more accurately report the program's net ex-post energy impacts. If all impact evaluations are required to report these metrics, California policy makers can use these estimates to update forecasts for the amount of energy that can be provided via energy efficiency, resource acquisition and procurement programs. These metrics are especially important now that California has committed to relying on these types of programs to provide a portion of the energy growth needs of the state's consumers.

   Minimum impact evaluation metrics should include:

   - First-year net kWh, kW, and therm impacts;
   - Net energy impacts for each consecutive year over which savings are expected (kWh, kW, and therms); and
   - Actual program expenditures (dollars actually spent to achieve program savings rather than dollars budgeted) less program evaluation costs.

   There is also a need to clearly define the term "*net*" so that it meets energy efficiency policy needs and is identical across all evaluations (including market effects studies that must estimate energy impacts from program-induced changes in the operations of energy markets). This may mean that the term "*net*" should be defined as savings adjusted for free-riders, participant spillover and persistence of effects.

5. **Require that all evaluations present the level of precision and the error bounds around estimates**. Have the evaluations contain a discussion of the reliability of the findings as well as the biases embedded within the approach or the analysis efforts and the approaches employed for reducing the influence of the biases.

6. **Provide clear guidance on how to conduct satisfaction assessments** so they lead directly to program improvement recommendations and can be compared across programs.

7. **Consider the use of market effects studies** to help quantify the market changes that are not captured within the program evaluation studies and the energy impacts associated with these indirect program effects. The evaluation field is moving toward assessments that incorporate moving market baseline estimates of technology adoption rates and causes, compared with the results from technology-specific or technology group-specific studies of total market changes, in order to identify how programs impact both the energy consumption of the participant and of the customers within the influenced market. California should continue to explore how these market-based evaluation approaches can be incorporated into the program evaluation process to provide more reliable short-term and long-term assessments of energy impacts.

8. **Clarify the relationship between program implementers and evaluators**. Do not establish evaluation relationships in which those responsible for program implementation and management are also responsible for evaluation budgeting and approach development efforts. There seemed to be misunderstanding for some implementers and evaluators about the term "arm's-length evaluation effort."
9. **Provide definitions of skills required of evaluation staffs**. This would be especially helpful for implementers seeking and assessing evaluators.
10. **Provide adequate time for the evaluation contracting and planning process**. Too often evaluation efforts were hindered by inadequate planning in these areas.
11. **Require evaluators to submit reports on time**. If programs are extended, consider whether evaluation reports are still needed within a reporting window so that portfolio assessments can move forward. Portfolio evaluation efforts were hindered by late or lacking reports.

The above-listed lessons learned and recommendations led the 2004-2005 Master Evaluation Contract Team to make recommendations for the 2004-2005 evaluation review process. As a result, the CPUC has used the 2002-2003 lessons to modify the 2004-2005 evaluation review process. Though it was not possible to incorporate all of the desired modifications, many have been incorporated and significant changes have already been undertaken within the Evaluation, Measurement and Evaluation (EM&V) Plans for the 2004-2005 program evaluations. However, the EEPM and administrator budget planning process still control the majority of the evaluation efforts and approaches.

Overview of the 2002-2003 Portfolio of Programs

From 2002 to 2003 California fielded a portfolio of eighty-five energy efficiency programs funded through its public goods charge (PGC) (2002-2003 Portfolio). Of these eighty-five programs, thirty-one were information- or education-focused programs, and the remainder represented a range of strategies to encourage adoption of energy-efficient technologies and practices, and are designed to directly acquire energy resources (resource acquisition programs).

Types of Programs Offered

The thirty-one information- or education-focused programs are officially classified as "information-only." Most of these programs and several of the resource acquisition programs also have educational components, however, the term "information-only" indicates that the CPUC did not require these programs to meet energy savings goals. All other programs are expected to set and meet established energy savings or resource acquisition goals. A presentation of the 2002-2003 Portfolio by these criteria is presented in Table 3. The individual programs are presented in Table 8.

Table 3. 2002-2003 Portfolio Program Classifications

Program Classification	Number of Programs	Percentage of Programs	Total Portfolio Expenditures	Percentage of Expenditures
Information-only	31	36%	$79,559,904	18%
Resource Acquisition	54	64%	$369,206,957	82%
Total Number of Programs	85	100%	$448,766,861	100%

The 2002-2003 Portfolio includes four basic types of programs characterized by their implementation agents. Three types of local programs were implemented through governments (Local-GOV), investor-owned utilities (Local-IOU) or third party implementers (Local-TPI). IOUs are also responsible for statewide (SW-IOU) programs. In 2002, the local and statewide IOU programs were funded for one year, while the non-utility programs were funded for a two-year period (2002-2003). To provide a comparison across programs during the 2002-2003 program cycle, program spending for 2003 (the second year of the IOU programs) is included to normalize the comparison with the non-utility programs. Table 3 and Table 4 reflect the distribution of programs for the 2002-2003 Portfolio (including the 2003 IOU programs). However, the 2003 statewide and local IOU program evaluations were not available for this report. Estimates of the evaluated savings for these programs were derived by applying the realization rates from the evaluation reports of the 2002 programs to the estimated savings projections for the 2003 programs.

Table 4. 2002-2003 Portfolio Program Types

Program Classification	Program Type	Percent of kWh Savings Goals	Percent of Natural Gas Savings Goals	Percent of Programs	Percent of Portfolio Expenditures
Resource Acquisition	Local-GOV	0%	1%	4%	1%
	Local-IOU	1%	7%	5%	2%
	Local-TPI	10%	28%	46%	15%
	SW-IOU	89%	64%	9%	65%
Information-Only	Local-GOV	N/A	N/A	5%	1%
	Local-IOU	N/A	N/A	14%	4%
	Local-TPI	N/A	N/A	11%	3%
	SW-IOU	N/A	N/A	7%	10%

The statewide IOU program savings and expenditures are the dominant feature of the 2002-2003 Portfolio. However, local programs implemented by third parties make up the largest number of programs. Given its size and decentralized nature, coordinating consistent evaluation within this group is challenging but critically important.

Targeted Market Sectors and Market Actors

Programs were engaged in a variety of market sectors including commercial, residential, local government, schools, industrial and agriculture. Several programs addressed multiple market sectors simultaneously through a crosscutting effort.

The largest proportion of information-only programs were crosscutting, targeting multiple markets. Most of these were statewide and local IOU programs. Most resource acquisition programs targeted residential and commercial markets, and were implemented by third parties. (Table 5, Table 6, Figure 1 and Figure 2)

Table 5. Targeted Market Sectors and Market Actors – Resource Acquisition Programs

Target Market	Target Market Actors	% of Budgets*	% RA Programs	% kWh Savings	% Therm Savings
Commercial: Small	Business Manager / Owner; Contractor / Builder; Manager: Plant, Facility, Property, etc.; Multi-family Property: Owner / Manager; Non-English-Speaking Business Manager / Owner; Renter; Retailer; Rural Business Owner / Manager; Waste Water Manager & Operator	7%	24%	4%	2%
Residential: All	Contractor / Builder; Distributor; Home Owner; Low Income Residence Owner; Manufacturer; Multi-family Property: Owner / Manager; Renter; Retailer	20%	11%	30%	3%
Crosscutting: Commercial Industrial	Business Manager / Owner; Contractor / Builder; Designer / Architect / Engineer; Distributor; Energy Manager; Manager: Plant, Facility, Property, etc.; Manufacturer; Non-English-Speaking Business Manager / Owner; Renter; Rural Business Owner / Manager	26%	6%	44%	28%
Residential: Single	CABEC Modeler; CHEERS Rater;	18%	13%	6%	26%

Target Market	Target Market Actors	% of Budgets*	% RA Programs	% kWh Savings	% Therm Savings
Family (SF)	Contractor / Builder; Home Inspector; Home Owner; Lender; Real Estate Professional; Retailer; Wholesaler				
Residential: Mobile Home	Home Owner; Multi-family Property: Owner / Manager; Non-English-Speaking Resident; Non-White Residents; Renter; Rural Resident	4%	6%	1%	6%
Commercial: Large	Business Manager / Owner; Community Leader; Contractor / Builder; Dealer; Designer / Architect / Engineer; Distributor; Energy Manager; Manager: Plant, Facility, Property, etc.; Wholesaler	14%	6%	9%	10%
Residential: Multi-Family (MF)	Business Manager / Owner; Multi-family Property: Owner / Manager; Non-English-Speaking Resident; Non-White Resident; Renter; Rural Resident	3%	6%	1%	8%
Crosscutting: All	Building Operator; Business Manager / Owner; Commercial Building Owner; Contractor / Builder; Home Owner; Multi-family Property: Owner / Manager; Renter; School Official / Administrator	2%	4%	0%	0%
Agriculture: Field Products	Business Manager / Owner; Grower Using Irrigation	1%	2%	0%	0%
Crosscutting: Community	Building Operator; Business Manager / Owner; Home Owner; Multi-family Property: Owner / Manager; School Official / Administrator; Students	1%	4%	1%	1%
Government: Other Local Government	Manager: Plant, Facility, Property, etc.	1%	2%	0%	0%
Industrial: All	Business Manager / Owner; Energy Manager; Manager: Plant, Facility, Property, etc.	1%	4%	1%	0%
Crosscutting: Residential Commercial	Business Manager / Owner; Contractor / Builder; Home Owner; Underserved Commercial Customer; Underserved Residential Customer	1%	2%	0%	0%
Government: Municipal	Waste Water Manager & Operator	1%	4%	0%	0%
Commercial: All	Business Manager / Owner; Dealer	1%	2%	0%	16%
Industrial: Small	Energy Manager; Manager: Plant, Facility, Property, etc.	0%	2%	1%	0%
Government: Water Agencies	Business Manager / Owner; Waste Water Manager & Operator	0%	1%	0%	0%
Institutional: Elementary Secondary Schools	School Official / Administrator	0%	2%	0%	0%
Agriculture: Animal Products	Dairy Farmer; Dealer	0%	2%	0%	0%

*Note: If programs were involved in more than one market, it was assumed that budgets were split equally. For example, if a program had two target markets half of the spending and energy savings was attributed to each market.

Table 6. Targeted Market Sectors and Market Actors – Information-only Programs

Target Market	Target Market Actors	% of Budgets*	% IO Programs
Commercial: Small	Business Manager / Owner; Manager: Plant, Facility, Property, etc.; Manufacturer	6%	8%
Residential: All	Home Owner; Non-English-Speaking Resident; Renter; Spanish-Speaking Resident	7%	10%
Crosscutting: Commercial Industrial	Building Operator; Business Manager / Owner; Energy Manager	19%	6%
Residential: SF	Contractor / Builder	2%	3%
Residential: MF	Multi-family: Owner / Manager	2%	3%
Crosscutting: All	Building Operators; Business Manager / Owner; Code Official; Contractor / Builder; Designer / Architect / Engineer; Education Center Attendee; Energy Manager; Home Owner; Manufacturer; Multi-family: Owner / Manager	31%	19%
Agriculture: Field Products	Business Manager / Owner; Contractor / Builder; Distributor; Grower Using Irrigation; Manager: Plant, Facility, Property, etc.; Waste Water Manager & Operator	2%	3%
Government: Other Local Government	County Government Leader; Local Government Official	2%	6%
Crosscutting: Residential Commercial	Building Operator; Business Manager / Owner; Chinese-Speaking Business Owner / Manager; Chinese-Speaking Resident; Code Official Community Leader; Contractor / Builder; Designer / Architect / Engineer; Energy Manager; Spanish-Speaking Resident; Home Owner; Korean-Speaking Resident; Local Government Official; Manager: Plant, Facility, Property, etc.; Manufacturer; Non-English-Speaking Business Manager / Owner; Vietnamese-Speaking Resident	20%	19%
Government: Municipal	Business Manager / Owner; Manager: Plant, Facility, Property, etc.	1%	1%
Government: Water Agencies	Business Manager / Owner; Contractor / Builder; Distributor; Grower Using Irrigation; Manager: Plant, Facility, Property, etc.; Waste Water Manager & Operator	2%	3%
Institutional: Elementary Secondary Schools	Building Operator; Business Manager / Owner; Designer / Architect / Engineer; Energy Manager; Manager: Plant, Facility, Property, etc.; Parent of Student; School Official / Administrator; School Teacher / instructor; Student	7%	17%

*Note: If programs were involved in more than one market, it was assumed that budgets were split equally. For example, if a program had two target markets half of the spending and energy savings was attributed to each market.

Figure 1. Distribution of Program Target Markets by Number of Programs

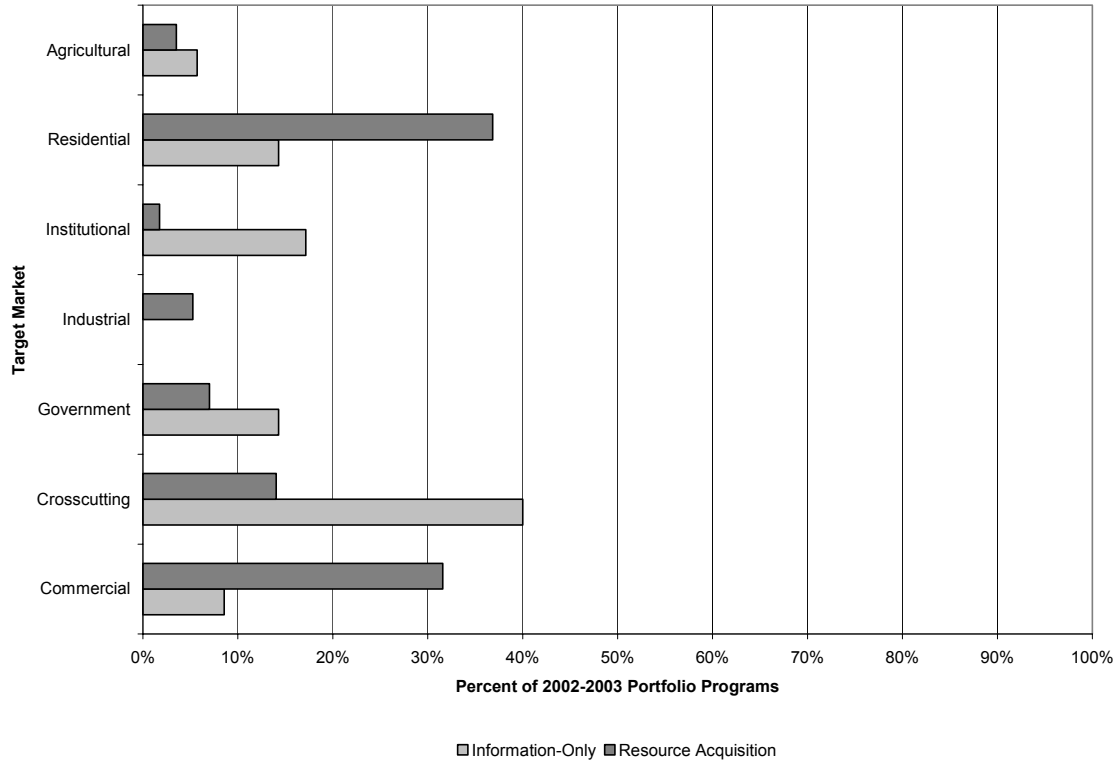
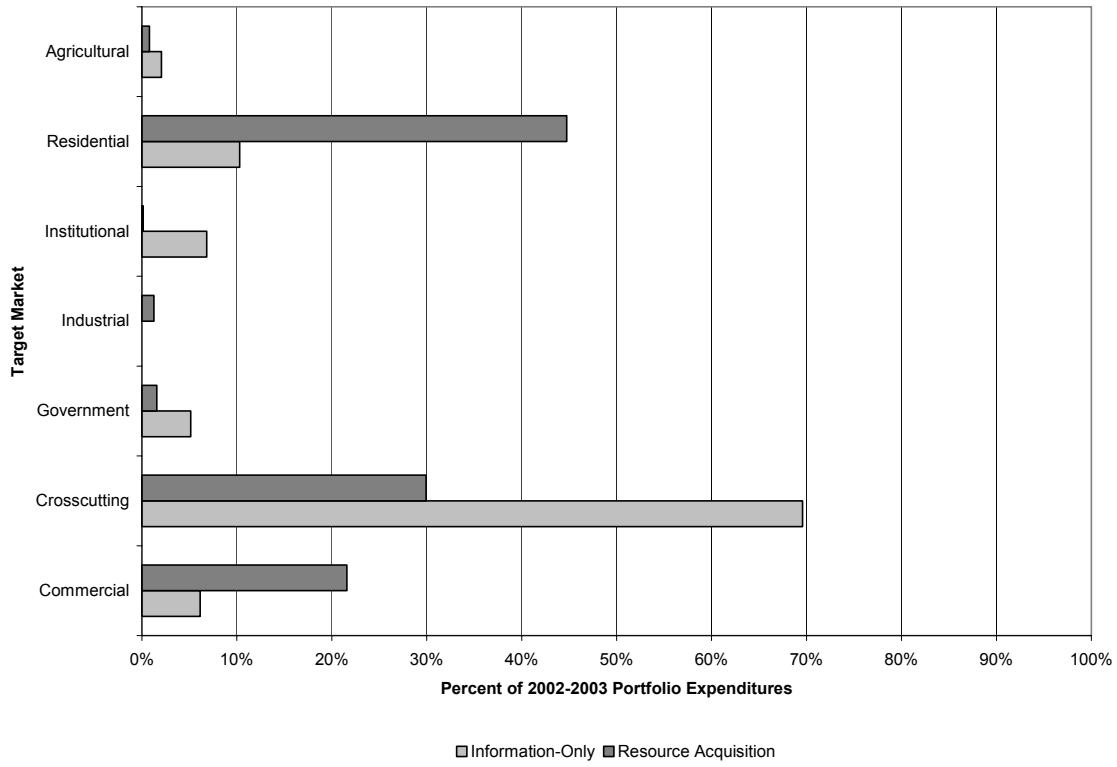


Figure 2. Distribution of Program Target Markets by 2002-2003 Portfolio Expenditures



Business managers and owners were the market actors most commonly targeted by both the resource acquisition and information-only programs. Assessment by proportion of budgets or energy savings achieved by these groups was not possible given the data available for this report. The distribution of targeted market actors by program type is presented in Table 7.

Table 7. Targeted Market Actors

Program Type	Actors Targeted*	Number of Programs	% of all RA Programs
Resource Acquisition	Business Manager / Owner	28	52%
	Home Owner	16	30%
	Contractor / Builder	14	26%
	Renter	11	20%
	Multi-family Property: Owner / Manager	10	19%
	Manager: Plant, Facility, Property, etc.	8	15%
	School Official / Administrator	4	7%
	Energy Manager	4	7%
	Distributor	4	7%
	Waste Water Manager & Operator	4	7%
	Non-English-Speaking Business Manager / Owner	4	7%
	Actors Targeted*	**Number of Programs**	**% of all IO Programs**
Information-Only	Business Manager / Owner	11	35%
	Building Operator	9	29%
	Contractor / Builder	8	26%
	Designer / Architect / Engineer	8	26%
	Home Owner	7	23%
	Manager: Plant, Facility, Property, etc.	6	19%
	School Official / Administrator	5	16%
	Energy Manager	5	16%
	Student	4	13%
	Renter	3	10%
	Manufacturer	3	10%
	Code Official	3	10%
	Local Government Official	3	10%
	Multi-family Property: Owner / Manager	2	6%
	Distributor	2	6%
	Waste Water Manager & Operator	2	6%
	Grower Using Irrigation	2	6%
	Chinese-Speaking Resident	2	6%
	School Teacher / Instructor	2	6%

*Note: The most common, but not all, targeted market actors are included in this table .

Each type of implementing party tended to focus its program efforts on different markets. For resource acquisition programs the local third party programs were concentrated in commercial and residential markets, statewide IOU programs in residential markets, local IOU programs in commercial markets and local government programs were evenly distributed across the various markets. IOUs generally employed crosscutting efforts for their information-only programs, while third parties targeted institutional markets, and local governments focused on government entities for the same. This information is presented graphically in Figure 3 and Figure 4.

Figure 3. Distribution of Program Target Markets by Program Type

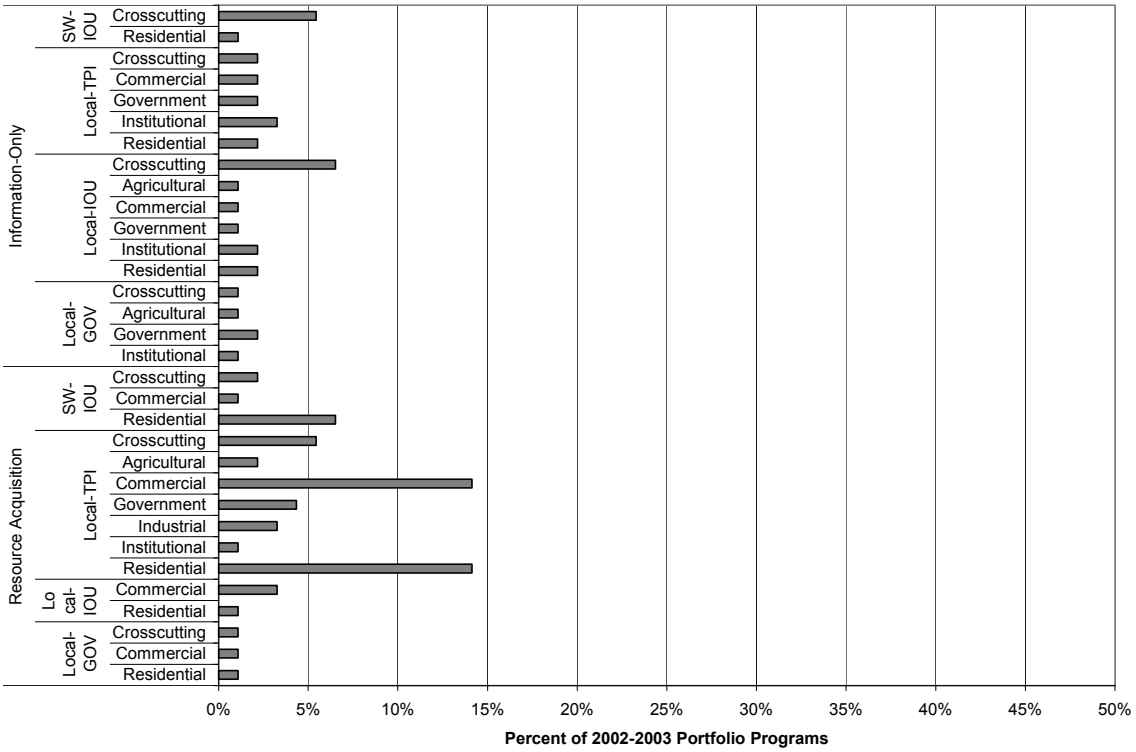
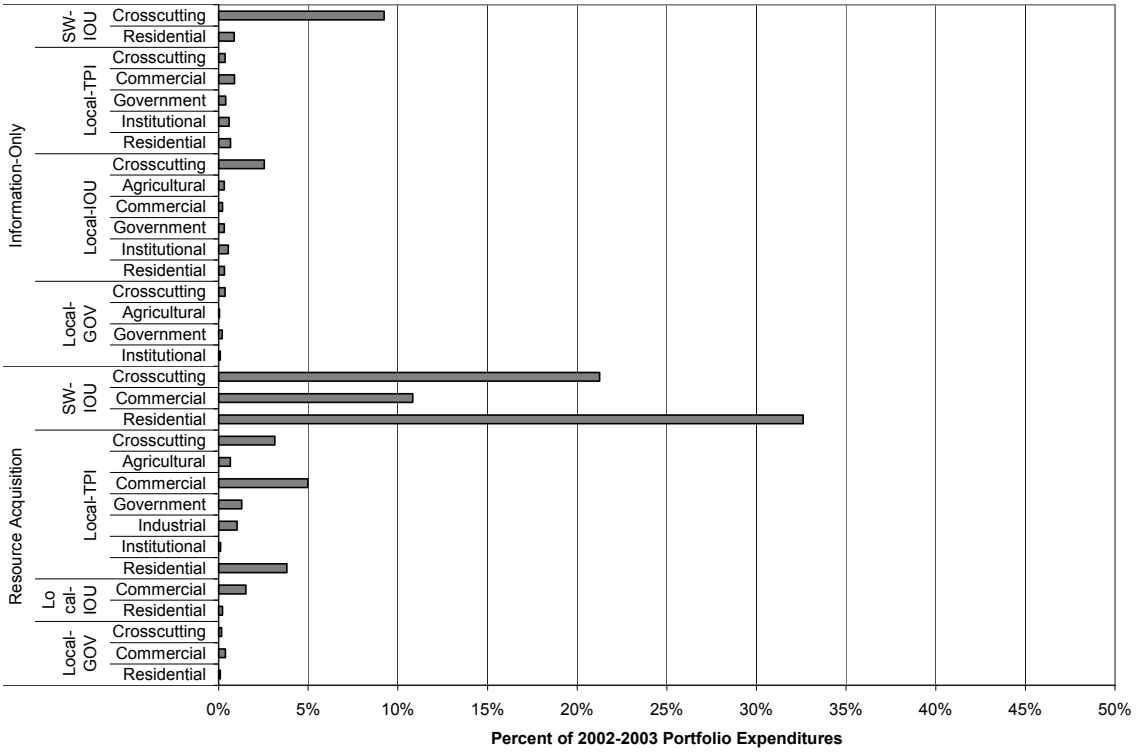


Figure 4. Distribution of Program Target Markets by 2002-2003 Portfolio Expenditures



Overview of Program Evaluation Efforts

The fifty-four resource acquisition programs and thirty-one information-only programs included in this report are presented in Table 8 along with evaluation-related information. Evaluation reports for five of these programs were not available at the time of writing this report.

Table 8. Evaluations of the 2002-2003 Portfolio

Program Name	Evaluation Contractor	Publication Date
Resource Acquisition Programs		
SW - IOU		
Appliance Recycling	KEMA-XENERGY	February 13, 2004
California ENERGY STAR® New Home Construction	RLW Analytics, Inc.	March 1, 2004
Cross-cutting Residential Lighting Program (Upstream Lighting)	KEMA-XENERGY; Quantum Consulting, Inc.	October 13, 2003
Energy Design Resources	Opinion Dynamics Corp.	December 2003
Express Efficiency	Quantum Consulting, Inc.	March 1, 2004
Multi-Family Rebates	Wirtshafter Associates, Inc.	February 27, 2004
Nonresidential Standard Performance Contract	Quantum Consulting, Inc.	Process / Market Assessment: March 25, 2004; Impact: November 18, 2004
Savings by Design	RLW Analytics, Inc.	July 2004
Single Family Home Energy Efficiency Rebate Program	Quantum Consulting, Inc.; KEMA-XENERGY	December 29, 2003
Local - IOU		
Hard To Reach Lighting Turn In	RLW Analytics, Inc.	November 15, 2003
Nonresidential Financial Incentives Program	RLW Analytics, Inc.	November 15, 2003
Small Nonresidential Hard to Reach Program	RLW Analytics, Inc.	November 15, 2003
Local EZ Turnkey Program	Quantec, LLC	July 22, 2003
Local - GOV		
San Diego Cool Communities Shade Tree Program	Zebedee & Associates	June 2004
San Diego Direct Install Small Commercial Program	Nexant	April 2004
Whole House Energy Retrofit	Business Economic Analysis and Research	June 30, 2004
Local - TPI		
Building Department and Small Builder Title 24 Standards Training	Sisson and Associates, Inc.	May 3, 2004
The Energy Savers Program	Sisson and Associates, Inc.	July 1, 2004
Comprehensive Hard-to-Reach Mobile Home Energy Savings	Robert Mowris & Associates	August 19, 2004
Comprehensive Hard-to-Reach Residential and Small Commercial Energy Savings Program**	Robert Mowris & Associates	October 29, 2004
Efficient Affordable Housing Program	Robert Mowris & Associates	August 18, 2004
Local Small Commercial Energy Efficiency & Market Transformation Program	Robert Mowris & Associates	June 30, 2004
Mobile Energy Clinic Program**	Robert Mowris & Associates	April 7, 2004
Proposal to Provide A Small Nonresidential Energy Fitness Program	Robert Mowris & Associates	April 30, 2004
The Gas-Only Multi-family Efficiency Program; "The Free Energy Project"	Robert Mowris & Associates	April 8, 2004
Time-of-Sale Home Inspection Program	Robert Mowris & Associates	November 18, 2004

Program Name	Evaluation Contractor	Publication Date
Upstream High Efficiency Gas Water Heater Program	Robert Mowris & Associates	October 19, 2004
Compressed Air Management Program	Ridge & Associates; The Draw Group; Equipoise Consulting, Inc.	January 7, 2005
The Oakland Energy Partnership Program		Not Complete - Program Extended
Energy Smart Grocer Program	Quantum Consulting, Inc.	March 25, 2004
Small Business Energy Efficiency Program (RightLights)	Quantec, LLC	June 3, 2005
Comprehensive Compressed Air Program	Quantec, LLC	July 7, 2004
Energy Efficiency Services for Electricity Consumption and Demand Reduction in Oil Production in the State of California	Quantec, LLC	June 30, 2004
Energy Efficient Local Government Program	Quantec, LLC	January 16, 2005
Energy Star CFL Program for Small Hardware and Grocery Retailers	Quantec, LLC	June 23, 2004
LiteVend	Quantec, LLC	June 30, 2004
Residential New Construction and Appliance Lighting Program	Quantec, LLC	April 20, 2004
South Bay Communities & Affiliates Energy Efficiency Program	Quantec, LLC	May 4, 2004
Stockton Area Comprehensive Local Program - Brighter Business	Quantec, LLC	June 18, 2004
Demand Control Ventilation Pilot Program	ICF Consulting	Not Complete -Program Extended
Municipal Wastewater Retro-Commissioning (PG&E Territory)	SBW Consulting	September, 2005
Municipal Wastewater Retro-Commissioning (SCE Territory)	SBW Consulting	September, 2005
Residential Duct Services Program		Not Complete
Check Me	Itron, Inc.	December 20, 2004
Electric & Gas Industries Association High-Efficiency HVAC Interest Rate Buy-Down Program	Summit Blue	August 2005
Davis Comprehensive Energy Efficiency Program (DCEEP)	Heschong Mahone Group	September 22, 2004
Agriculture Pumping Efficiency Program	Equipoise Consulting Inc.; Ridge & Associates; Vanward Consulting; California AgQuest Consulting Inc.	June 17, 2004
LightWash	Equipoise Consulting (Washing Machines) ; Quantum (Lighting)	June 30, 2004
Mobile Home Energy Efficiency and Education Program	Business Economic Analysis and Research	June 30, 2004
California State University Chancellor's Office 2002--2003 Energy Efficiency Program	Aloha Systems	May 5, 2004
The County of Los Angeles Internal Services Division Energy Efficiency Program	Aloha Systems	May 20, 2004
The Energy District Approach for Sustainable Energy Efficiency in California	Aloha Systems	July 13, 2004
California Variable Speed Drive Farm Program	kW Engineering	June 8, 2005
Pre-Rinse Spray Head Installation for the Food Service Industry		Not Complete
Information-only Programs		
Statewide - IOU		
Home Energy Efficiency Surveys	Ridge & Associates; KVD Research Consulting; Quantum Consulting, Inc.	June 1, 2004
Emerging Technologies	Ridge & Associates; Equipoise Consulting, Inc.	December 26, 2003
Building Operator Certification and Training	Research Into Action, Inc.	November 30, 2003

Program Name	Evaluation Contractor	Publication Date
Audit Program	Quantum Consulting, Inc.	March 1, 2004
Education & Training	KEMA-XENERGY	December 16, 2003
Statewide Codes & Standards	ADM Associates, Inc.	June 2004
Local - IOU		
Local Government Initiative	Wirtshafter Associates, Inc.	November 20, 2004
Diverse Markets Outreach Program	RLW Analytics, Inc.	December 8, 2003
In Home Audits	RLW Analytics, Inc.	November 15, 2003
Local Energy Code Training	RLW Analytics, Inc.	November 15, 2003
Energenius	Ridge & Associates; Equipoise Consulting, Inc.; Vanward Consulting	December 27, 2003
School Resources Program	Ridge & Associates; Equipoise Consulting, Inc.; Vanward Consulting	December 27, 2003
Local Small Business Energy Assessment (SBEA)	Quantec, LLC	July 22, 2003
Pacific Energy Center (PEC)	Newcomb Anderson Associates	December 2003
Residential In-Home Energy Survey Program	KVD Research Consulting	April 26, 2004
Pump Tests & Hydraulic Services Program	Equipoise Consulting Inc,; Ridge & Associates; Vanward Consulting; California AgQuest Consulting Inc.	November 14, 2003
Demonstration & Information Transfer	Energy Market Innovations, Inc.	February 26, 2004
Local Codes & Standards Program		Not Complete
Local - GOV		
San Diego K-12 Energy Education Program	Zebedee & Associates	June 2004
San Diego Public Agency Information and Technical Support Program	Zebedee & Associates	June 2004
San Diego Region Agriculture, Water and Energy Program	Zebedee & Associates	April 2004
San Diego Region Energy Resource and Education Center	Zebedee & Associates	June 2004
Local - TPI		
Green Schools, Green Communities	Vanward Consulting; Equipoise Consulting, Inc.; Ridge & Associates	July 9, 2004
School Energy Efficiency Program	Vanward Consulting; Equipoise Consulting Inc.; KEMA-XENERGY .; Ridge & Associates; Shel Feldman Management Consulting	February 14, 2005
Green Building Technical Support Services	Shel Feldman Management Consulting	March 2005
Comprehensive Whole-House Residential Retrofit Program	Lutzenhiser & Associates	August 15, 2004
Partnership for Energy Affordability in Multi-Family Housing	KEMA-XENERGY	June 23, 2004
GHPC's Program To Promote Geoexchange	Itron, Inc.	May 2004
Regional Energy Authority Pilot Projects	Heschong Mahone Group	July 7, 2004
Energy Efficiency in Commercial Food Service	Equipoise Consulting, Inc.; Quantum Consulting Inc.; Energy Solutions; RJ Research	April 2, 2004.
Chinese Language Efficiency Outreach (CLEO)	ASW Engineering Management Consultants, Inc.	December 10, 2003

The focus of evaluation efforts varied across programs. This variation was primarily driven by program focus, implementer, evaluation desires and CPUC-approved evaluation objectives. However, the rigor and comprehensiveness of the evaluation effort was determined in most cases by the available supportive budget. Both information-only and resource acquisition program evaluations included process evaluation components. As expected, resource acquisition program evaluations more frequently included impact evaluations, although a few information-only program evaluations also did. (Figure 5).

Figure 5. Evaluation Types Included in Evaluation Reports (by Percent of Programs)

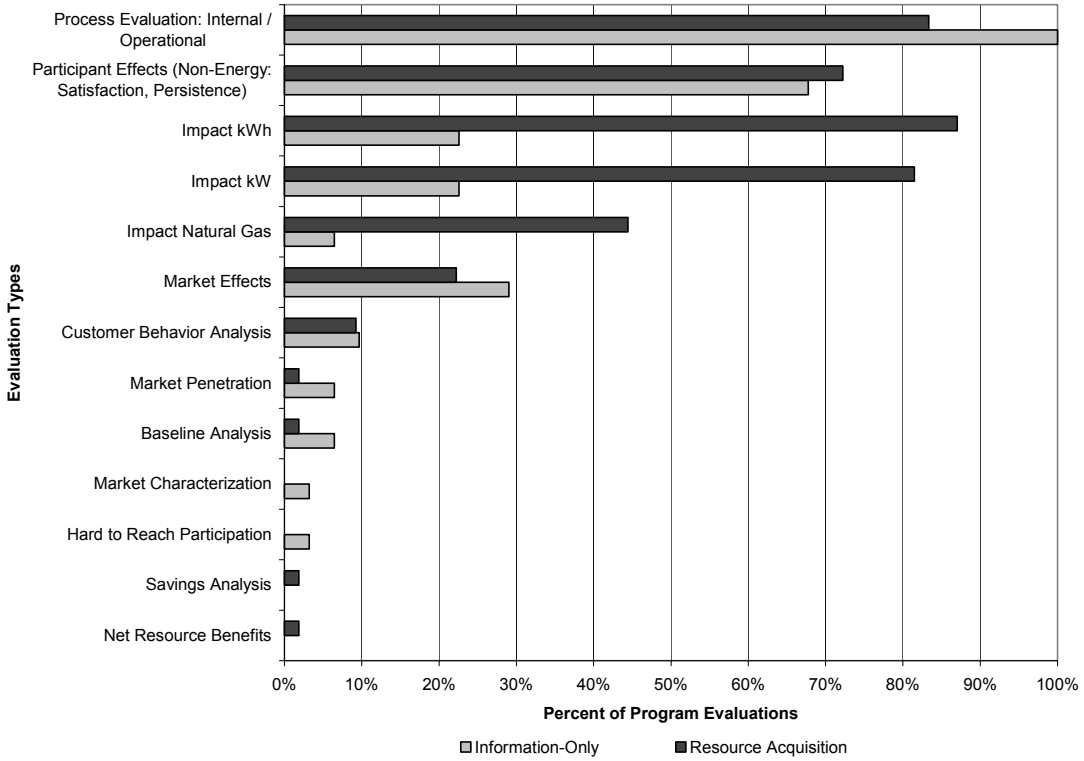
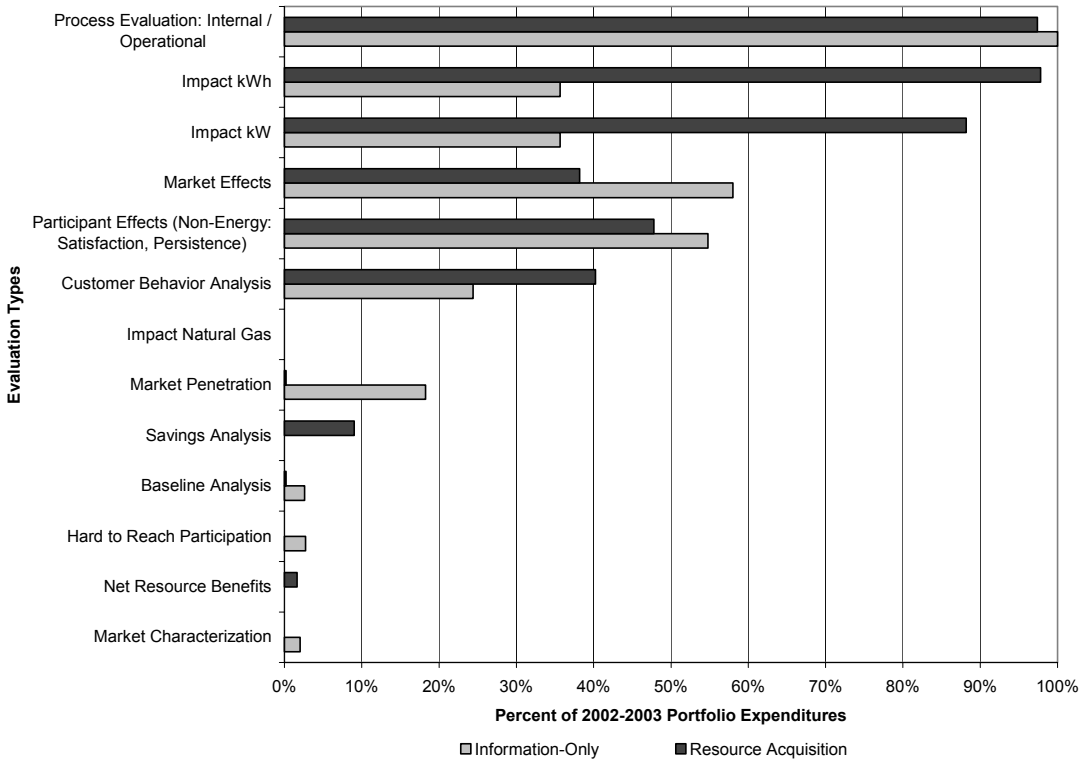


Figure 6. Evaluation Types Included in Evaluation Reports (by Percent of 2002-2003 Portfolio Expenditure)



As expected, the evaluation methods employed are closely related to the type of evaluation conducted. Understandably, qualitative evaluation methods were more prevalent in assessing information-only programs, while resource acquisition program evaluations relied more heavily on quantitative methods to gauge energy savings.

As depicted in Figure 7, both resource acquisition and information-only program evaluations relied heavily on participant surveys, but resource acquisition program evaluations were more apt to use on-site inspections and audits to evaluate energy savings. Most (70 percent) of the resource acquisition program evaluations, accounting for 55 percent of expected kWh savings, in some way complied with the EEPM requirement to use IPMVP evaluation approaches (which call for on-site measurements). However, many evaluations did not conduct on-site or site-specific verification, monitoring or metering activities. Program staff interviews were also used in information-only program evaluations.

Figure 7. Evaluation Methods by Number of Program Evaluations

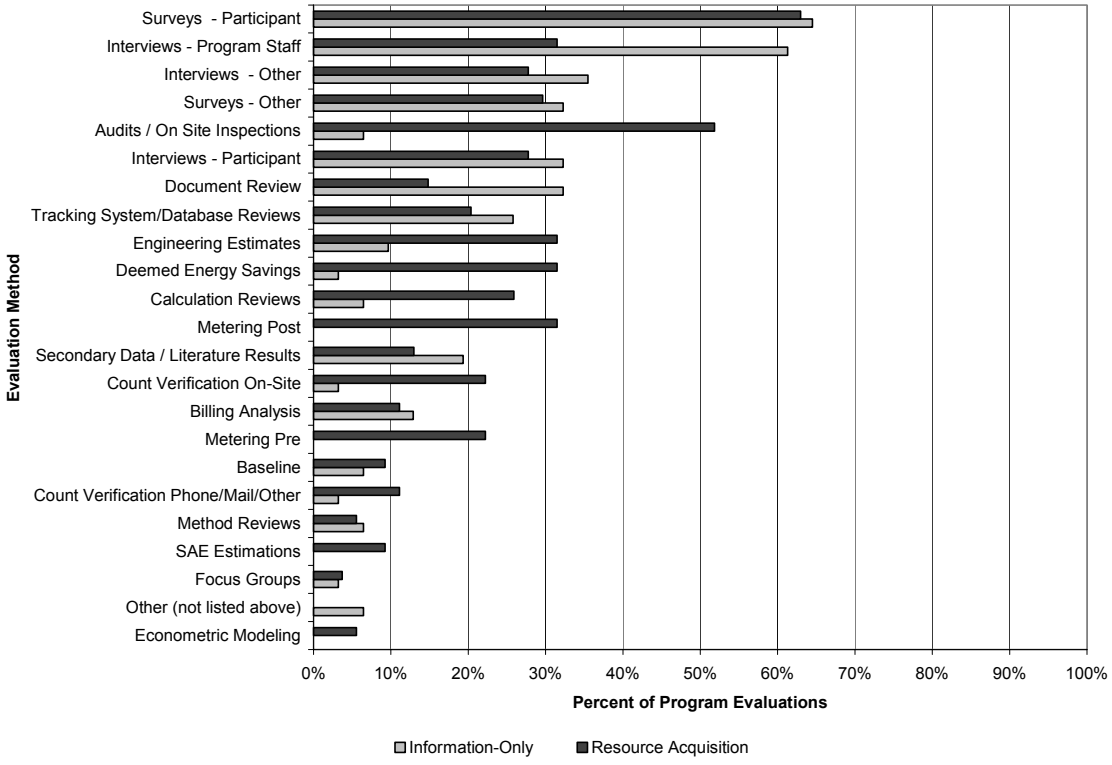
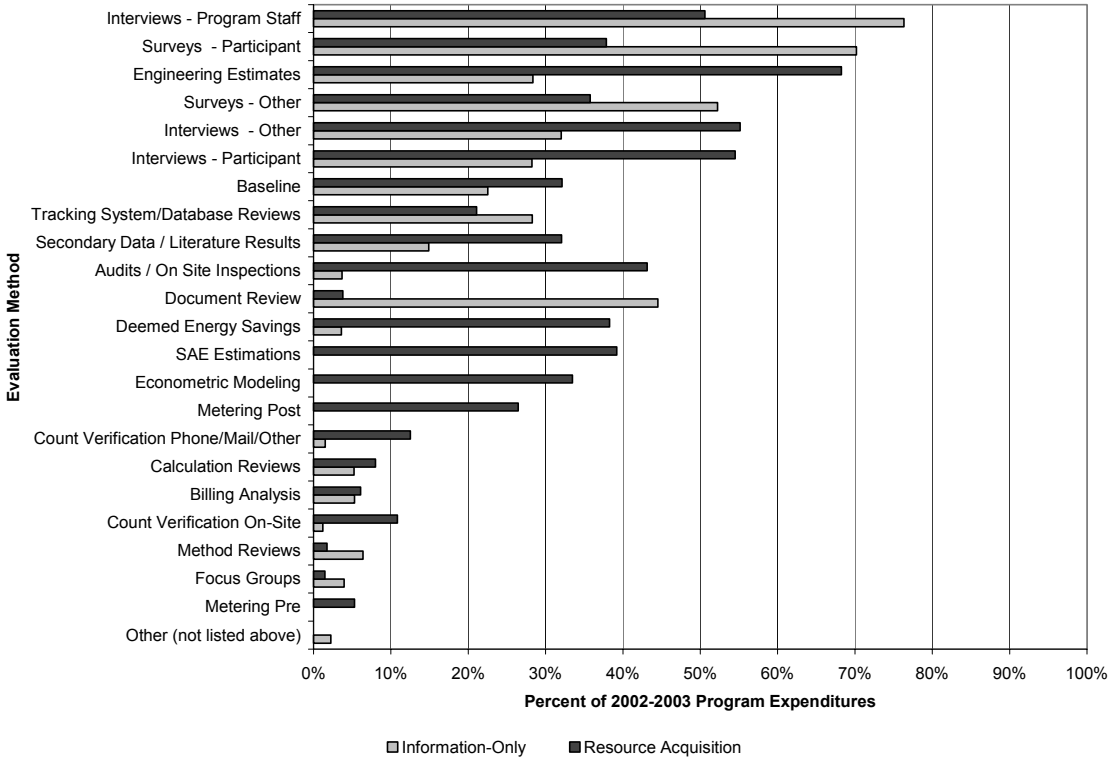


Figure 8. Evaluation Methods by Evaluated Program Expenditures



Impact Evaluation Efforts

Impact evaluations are critical to the successful evaluation of California's energy program portfolio. Not only do these studies document the savings Californians receive for their funding, they also identify specific program resources and related costs. In addition, impact evaluations serve to determine if deemed savings estimates need to be revised by comparing them to achieved savings.

Accurate measurement of energy resources acquired through energy efficiency programs is essential if these resources are to be considered for policy planning to meet California's energy needs. The *California Evaluation Framework* (*Evaluation Framework*) offers a good summary program impact evaluation expectations:

> The primary purposes of conducting a summative impact evaluation are (1) to provide reliable estimates of the energy and demand savings created by the program(s) for use in cost-effectiveness analysis, (2) to know how much resource can be depended upon, and (3) to incorporate savings in overall Public Goods Charge (PGC) fund estimates. In all of these cases what is desired is the energy and demand savings induced by the program.[4]

[4] TecMarket Works, *The California Evaluation Framework* (Southern California Edison Company, 2004), Study ID K2033910, 133. Can be obtained at: http://www.calmac.org/search.asp. Enter "California Evaluation Framework" and download the 500-page reference document as an Adobe .pdf file.

The three primary components of energy and demand savings are:

1. Measures/actions installed/adopted;
2. Savings achieved through those changes; and
3. Savings over what would have occurred without the program.

These savings need to be able to be assessed for reliability and incorporated in the overall savings estimates.

Programs achieve energy and demand savings by participants installing or adopting measures (or, in the case of market transformation efforts, driving others to do so). At a minimum, impact evaluations must provide savings estimates obtained either through proper analysis of consumption (utility metering data) or verification of the installations/actions and the resulting savings.

In general, the impact evaluations of the 2002-2003 Portfolio programs provided estimates of annual energy savings and, in some cases, demand impacts, but did not meet the basic criteria necessary for a reliable evaluation in all cases. Specifically, the rigor and precision of the impact evaluations for many studies are not of the caliber to provide reliable estimates of the actual contributions of the 2002-2003 Portfolio to meeting the state's energy needs. This section of the report presents findings from the evaluation reports and discusses the associated evaluation methods and rigor.

The analysis in this section is based on a review of the 2002 IOU program evaluations and the 2002-2003 non-IOU program evaluations. As of the writing of this report, not all of the 2003 IOU program evaluations are complete. Since evaluations of similar 2002 IOU programs were available, the matching 2003 evaluated program savings were estimated based on the achievements of the 2002 programs. (For program year (PY) 2002, local and statewide IOU programs were funded for one year, while the non-utility programs were funded for a two-year period (PY 2002-2003)). In order to provide a picture of the 2002-2003 Portfolio as a whole over the two-year timeframe, in tables with evaluated energy savings, the realization rates[5] from the available evaluations were applied to 2003 ex-ante savings reported in the final workbooks for the IOU programs.

Evaluated Savings Obtained versus Anticipated Savings

Annual Energy and Demand Savings

Based on resource acquisition program implementation plan savings targets,[6] the 2002-2003 Portfolio was anticipated to save 2,164,315,460 kWh and 46,498,851 therms. Of the originally anticipated savings, over 99 percent of the kWh and 80 percent of the therm saving were evaluated to some degree in reports completed as of the writing of this report. At this time, five studies have not been reviewed for the 2002-2003 programs. Energy savings for these programs

[5] Realization rates were calculated by dividing the evaluated savings for the 2002 program by the claimed savings in the final workbook for 2002. For some programs the realization rate was taken directly from the evaluation report.

[6] Taken from the program final workbooks.

are not included in the related discussion. Four information-only program evaluations also reported energy savings. These programs are not included in the following graphics, but are reviewed in the following section of the report.

In reviewing the evaluation reports it was assumed that, unless specified differently, the reported evaluation savings are based on the final program-claimed savings and installation counts in the final workbooks for each program.[7] Evaluated savings are the most rigorously measured savings in the evaluation reports. In some cases the program-claimed savings covered in the evaluation report are not the same as the program-reported savings from the final workbooks. This occurs for various reasons. For example, the evaluation may cover a different timeframe than the program-reported savings or the list of measures reported by the programs may not have all been evaluated.

Each evaluation was unique in the method used to report evaluated savings and adjustments applied to reported savings. To develop an accurate comparison of the energy savings goals, the savings claimed, and the evaluated savings, only programs using these metrics could be included in the comparison. This process of elimination resulted in only a portion of the total 2002-2003 Portfolio goals being compared. In this comparison, the program-claimed savings (from the final workbooks) and the evaluated savings were divided to develop a program-specific realization rate. With the assumption that evaluated savings were based on program-claimed savings, programs that had a realization rate greater than 150 percent or less than 50 percent of the program-reported savings were double-checked with their evaluation reports. If the realization rate in the evaluation report was within 6 percent of the calculated realization rate the program is included in the following comparison table. If the difference between the rates was greater than 6 percent, the program was not. The statewide and local IOU programs were the exception to this process. For these programs the realization rates reported in the 2002 program studies were applied to the matching 2003 programs to develop an estimate of energy savings for both 2002 and 2003.

Of the 2002-2003 Portfolio energy savings goals evaluated, only 56 percent of kWh savings, 32 percent of therm savings, and 47 percent of kW savings could be included in this comparison. The following tables present the results of this comparison.

The total 3,383,814 kWh savings goal for the five programs without evaluation reports at the time of this writing represent less than 1 percent of that for the 2002-2003 Portfolio as a whole. The savings goal of 2,160,931,646 kWh evaluated in the available reports represents 99.8 percent of the overall 2002-2003 Portfolio kWh savings goal. However, only 56 percent of these savings could be accurately compared with program goals and claimed savings.

[7] Reported on worksheet "T10 - Annual Report Summary."

Table 9. kWh Savings Goals, Claimed and Evaluated

(Represents 56% of the Evaluated Savings Goal)

Type	kWh Goals	Claimed kWh	Evaluation Estimated kWh Savings	Percent Claimed vs. Goals	Evaluated versus Claimed Realization Rate	Evaluated to Goal Achievement Rate
Local-GOV	4,491,000	5,341,667	4,280,328	119%	80%	95%
Local-IOU*	15,828,201	8,357,868	13,060,337	53%	156%	83%
Local-TPI	137,647,786	138,514,795	118,547,460	101%	86%	86%
SW-IOU*	1,062,018,741	1,285,409,084	1,134,214,131	121%	88%	107%
Total	**1,219,985,728**	**1,437,623,414**	**1,270,102,256**	**118%**	**88%**	**104%**

*Note: Energy savings goals and claimed savings for both Program Year (PY) 02 and PY03 are included in this calculation. The evaluated results were calculated based on the realization rates achieved in PY02 and reported in those evaluation reports.

At the portfolio level, for the sub-set of programs included in the comparison, the overall program-reported savings indicates that the programs achieved 118 percent of their goals with all programs reporting achievements greater than goals. However, only 88 percent of the claimed savings could be documented by evaluations, establishing an evaluation-confirmed realization rate of 88 percent. Applying this rate to reported savings indicates that the 2002-2003 Portfolio achieved 104 percent of its goals.

The total 7,766,372 therm savings goals for the five programs without available evaluation reports represents 20 percent of the total 2002-2003 Portfolio goal. The savings goal of 38,732,478 therms evaluated represents 80 percent of the overall 2002-2003 Portfolio therm savings goal. However, only 32 percent of this goal (12,425,064 therms) could be accurately compared with program goals and claimed savings (Table 10).

Table 10. Therm Savings Goals, Claimed and Evaluated

(Based on 32% of the Evaluated Savings Goal)

Type	Therm Goals	Claimed Therms	Evaluation Estimated Therms Savings	Percent Claimed vs. Goals	Evaluated versus Claimed Realization Rate	Evaluated to Goal Achievement Rate
Local-GOV	0	0	0	0%	0%	0%
Local-IOU	0	0	0	0%	0%	0%
Local-TPI	2,591,778	2,208,840	1,904,732	85%	86%	73%
SW-IOU	9,833,286	9,946,713	8,207,781	101%	83%	83%
Total	**12,425,064**	**12,155,553**	**10,112,513**	**98%**	**83%**	**81%**

At the portfolio level, for the sub-set of programs included in this comparison, the overall program-reported savings indicate that the programs achieved 98 percent of therm savings goals. The evaluation efforts were only able to document 83 percent of the claimed savings, establishing an evaluation-confirmed realization rate of 83 percent. Applying this rate to reported savings indicates that the 2002-2003 Portfolio achieved 81 percent of its natural gas savings goals.

The total 2,982 kW demand savings goals for the five programs without available evaluation results represents 1 percent of the total 2002-2003 Portfolio goal. The demand savings goal of

499,611 evaluated represents 99 percent of the overall 2002-2003 Portfolio demand savings goal. However, only 47 percent of these savings could be accurately compared with the program goals and claimed savings.

Table 11. Demand Savings Goals, Claimed and Evaluated

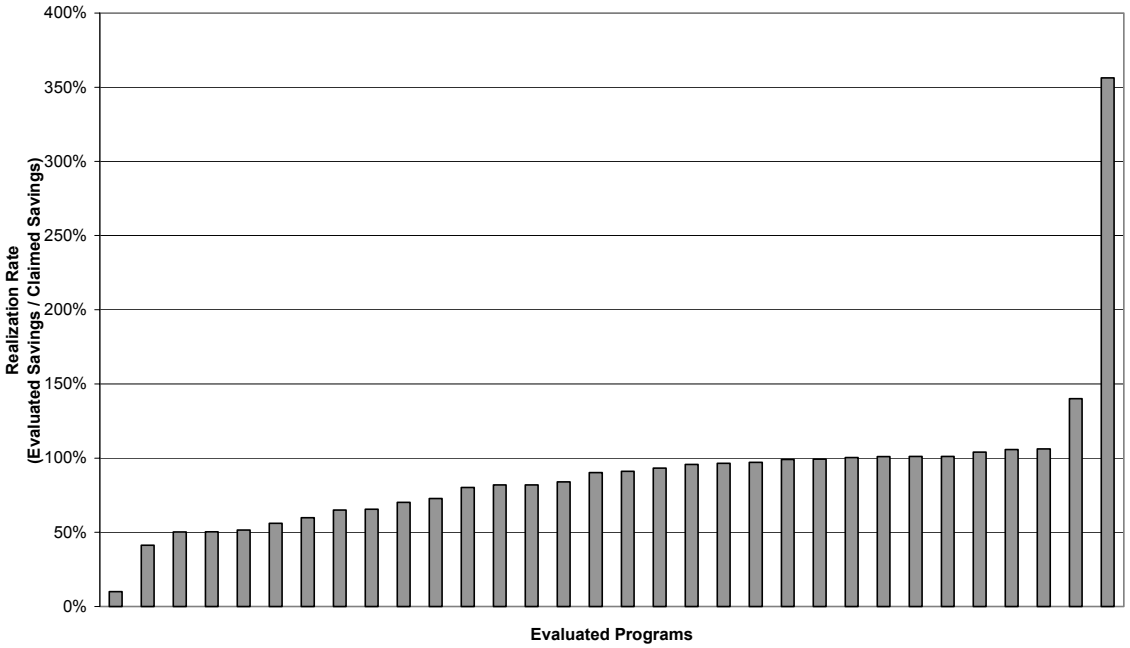
(Based on 47% of Evaluated Savings Goal)

Type	Demand Goals	Claimed Demand Savings	Evaluation Estimated Demand Savings	Percent Claimed vs. Goals	Evaluated versus Claimed Realization Rate	Evaluated to Goal Achievement Rate
Local-GOV	1,219	1,361	1,319	112%	97%	108%
Local-IOU	993	817	827	82%	101%	83%
Local-TPI	32,496	35,346	23,701	109%	67%	73%
SW-IOU	199,433	206,045	174,872	103%	85%	88%
Total	**234,140**	**243,569**	**200,719**	**104%**	**82%**	**86%**

At the portfolio level, for the programs that could be compared in this analysis, administrators claimed to achieve 104 percent of their goals. At the implementation category summary level, all of these programs, except Local IOU programs, claimed to exceed their goals. However, the realization rate, based only on the evaluated savings for these programs was 82 percent at the portfolio level. The evaluation-confirmed goal achievement rate for the 2002-2003 Portfolio was 86 percent.

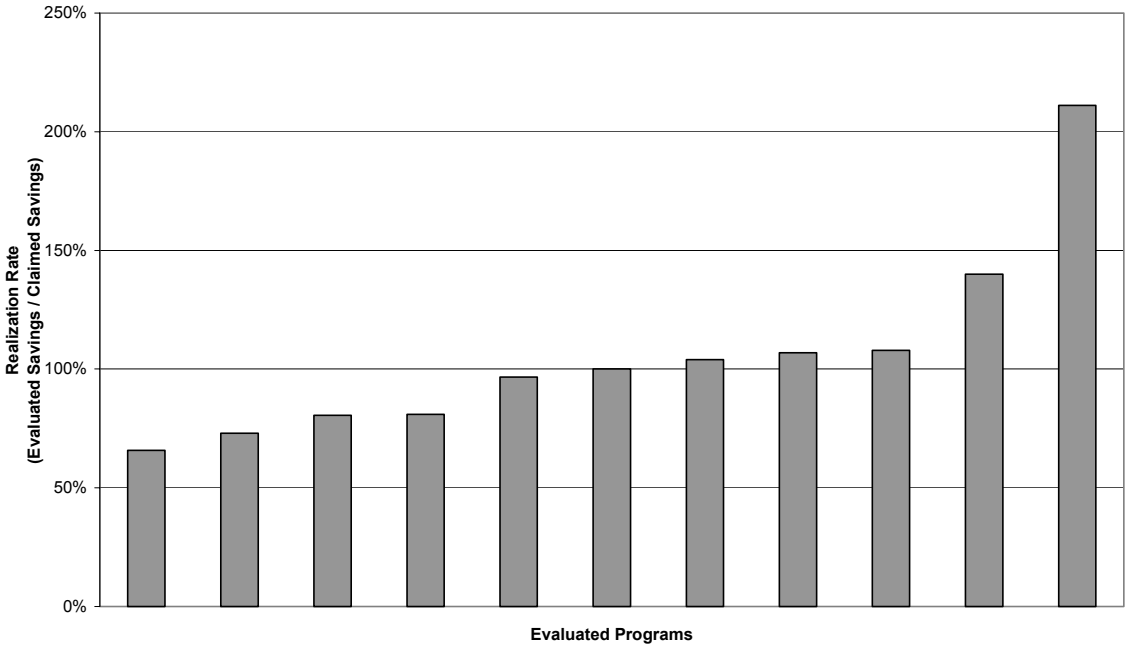
In Figure 9, Figure 10, and Figure 11 the evaluation confirmed realization rate by program is represented graphically.

Figure 9. kWh Energy Savings Realization Rate by Program



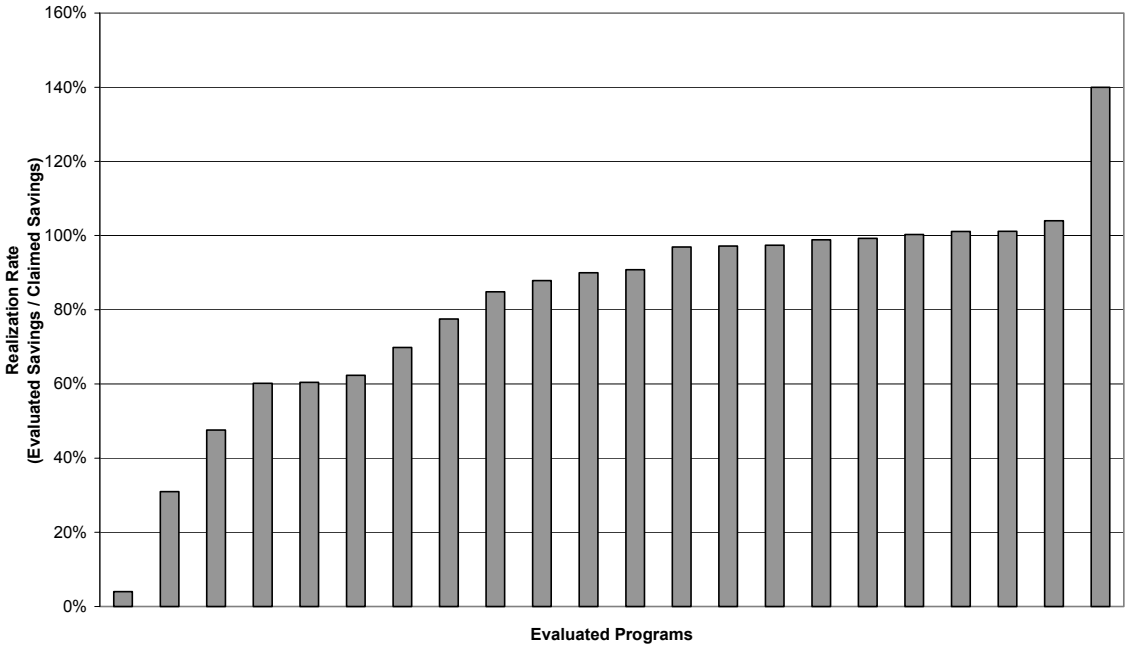
The average and median realization rates for kWh savings were both 91 percent. One local third party program (CHEERS Local ENERGY STAR Homes Program) was evaluated to have a realization rate of 349 percent because homes that were built as part of the program were considerably more efficient than expected.

Figure 10. Therm Energy Savings Realization Rate by Program



The average realization rate for therm savings was 106 percent and the median realization rate was 100 percent. One local third party program (CHEERS Local ENERGY STAR Homes Program) was evaluated to have a realization rate of 211 percent because homes that were built as part of the program were considerably more efficient than expected.

Figure 11. Demand Savings Realization Rate by Program



The average realization rate for kW savings was 80 percent and the median realization rate was 90 percent. One program was evaluated to have a 140 percent realization rate when participant and non-participant spillover were considered. Non-participant spillover is not reported in most of the other studies.

Lifetime Energy and Demand Savings

Although an important measure of the impact of programs over time, lifetime energy savings are seldom examined in the evaluations of 2002-2003 Portfolio programs, and annual load impacts are not reported in any of them. Only 12 of 50 (24 percent) resource acquisition program evaluations reported lifetime energy savings, reflecting about 2 percent of the 2002-2003 Portfolio kWh savings goal and about 8 percent of the therm savings goals. This level of reporting is insufficient to draw conclusions about the contributions of these programs to the long-term energy needs of the state and demonstrates the need for prescriptive evaluation protocols and reporting requirements.

Impact Evaluation Methods Used, Completeness and Summary of Rigor Assessment

Equally or more important than the amount of energy savings reported is the accuracy and reliability of the claimed savings. The veracity of savings claims are not dependable if estimates are unreliable, biased or substantially inaccurate.

This section first provides an overview of the methods that were used in the impact evaluation studies, then reviews the prevalence of various reliability and comparability related components including their application of:

- Net-to-gross Analysis;
- Installation Verification and Consumption Analysis;
- Free-ridership;
- Positive Adjustments for Spillover;
- Precision; and
- Cost-effectiveness

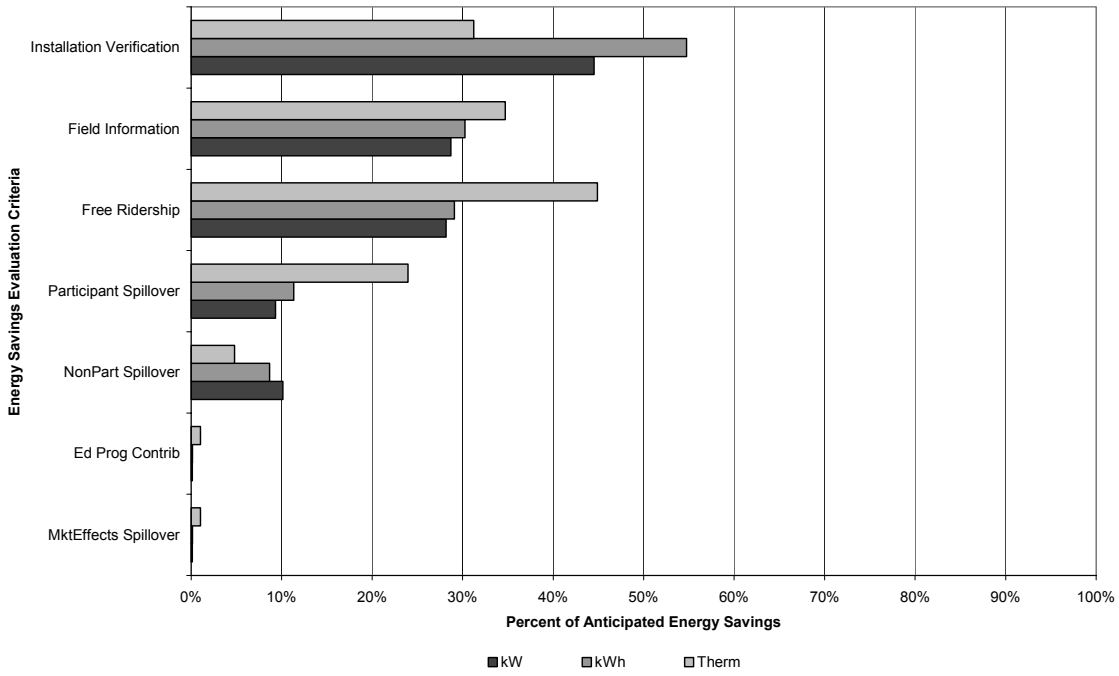
Overview of Impact Evaluation Methods Used

Impact evaluations can use a variety of methods within the associated M&V and calculation approaches used to estimate energy savings gained from programs. Not all methods for reviewing measure-level or program-level energy savings provide equal or comparable rigor. The most reliable estimation approaches are engineering-based methods that use metered data linked with consumption-savings modeling, or employ well thought-through engineering estimation approaches linked with on-site examination of technology application and use conditions. Employing rigorous regression analysis approaches using consumption data (utility metered data) is also reliable. Approaches using reviews of engineering algorithms, simply applying deemed savings estimates or using deemed savings adjusted from survey or interview collected data are far less so. The level of accuracy of these latter methods depends on how well program operations match those of prior studies employing rigorous approaches (if used) or on the accuracy of the assumed operational conditions (baseline and technology use) when used to adjust deemed estimates.

Program administrators and evaluators were instructed to use EEPM evaluation guidelines in preparing evaluation budgets and plans. The EEPM requires the use of the IPMVP or a discussion of planned M&V efforts with an explanation of why the IPMVP cannot be used. IPMVP methodology rigor depends on applying IPMVP approaches to the savings components with the greatest uncertainty and applying metering, measurement and monitoring techniques to lessen the uncertainty of the estimates applied to these components. Similarly, regression analysis of pre- and post-retrofit billing is dependent on a variety of quality factors in the data preparation, methods, applications and analysis. 2002-2003 Portfolio program evaluators did not consistently apply the most rigorous methods for evaluating energy savings impacts making many of the conclusions incomparable across studies.

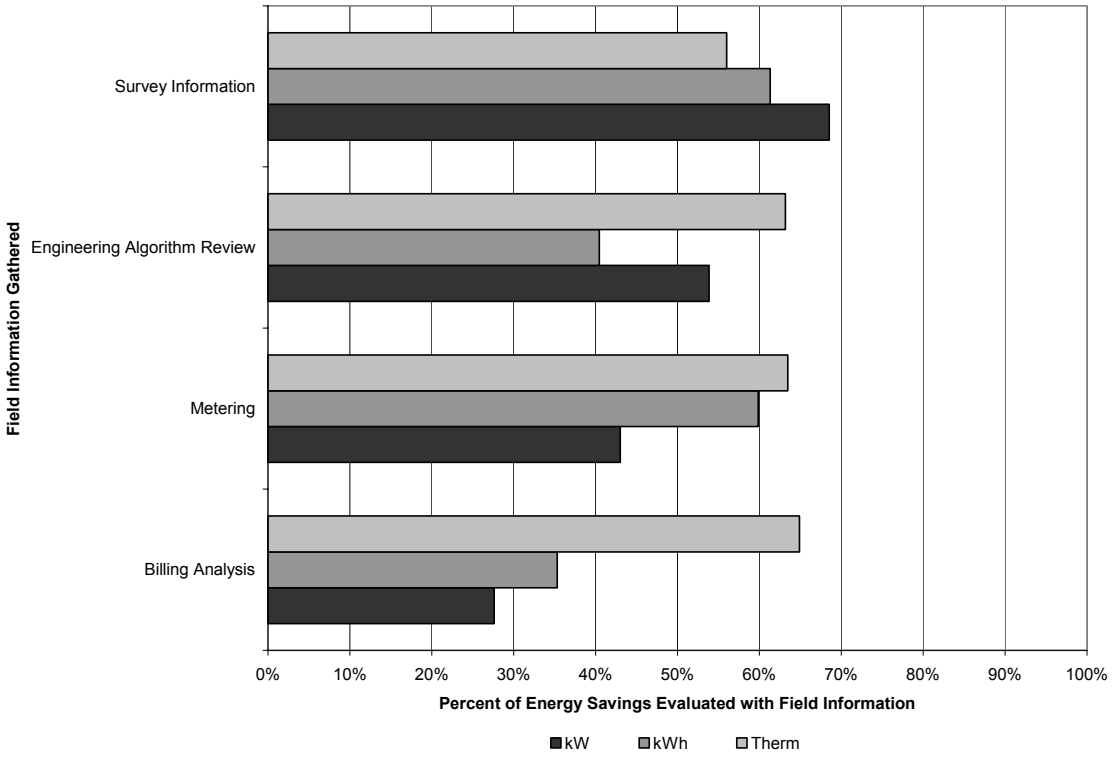
The different types of methods used to evaluate the energy savings impacts of the 2002-2003 Portfolio are presented in Figure 12. Note that only 55 percent of the anticipated kWh savings were evaluated using verification of installation approaches, even though IPMVP on-site field protocols were required of all evaluations. Less than half of the therm savings and demand savings were evaluated using verification of installation approaches. Non-verification field information of various types was also used to evaluate and adjust 30 percent of the anticipated kWh savings, slightly more for therm savings and slightly less for demand savings. Almost 30 percent of the anticipated kWh and demand savings were adjusted for free-ridership. Nearly 45 percent of therm savings were evaluated for free-ridership. Of the portfolio demand and kWh savings 11 percent were adjusted for participant spillover; roughly 25 percent of the therm savings were adjusted for participant spillover. Less than 10 percent of either demand, kWh or therm savings were adjusted for non-participant spillover. Less than 1 percent of the demand, kWh and therm savings were adjusted for education or market effects. It is clear from these findings that the program impacts presented in the evaluation reports are not comparable across the studies conducted. This, however, doesn't necessarily mean that the reported energy savings estimates are wrong, only that it is impossible to draw conclusions about the level of accuracy of a significant portion of them.

Figure 12. Energy Impact Evaluation Methods Used



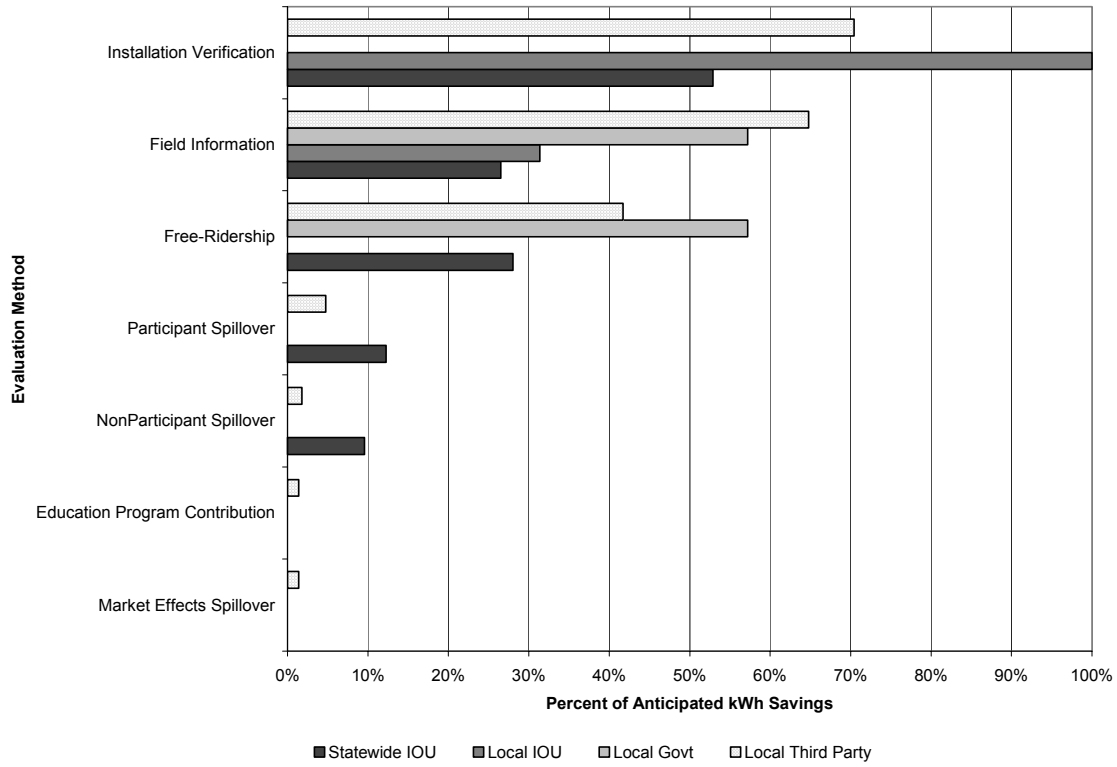
A variety of tools were used to gather information from the field to support energy savings estimates. Figure 13 presents the most commonly used of these based on energy savings. Surveys and metering were used most frequently for estimating kWh savings followed by engineering algorithm reviews and billing analysis approaches. The largest portion of therm savings was evaluated using billing analysis followed by metering and engineering algorithm reviews. Survey information was the least common data collection tool for therm savings. Survey information was the most common field method to assess demand savings followed by engineering algorithm review, metering and billing analysis.

Figure 13. Field Methods Used to Adjust Energy Savings



A closer look at methods used to evaluate kWh savings reveals that all of the local IOU program evaluations used installation verification methods, while none of the local government program evaluations did so (Figure 14). Over half of the latter used other field-collected information and free-ridership adjustments. The evaluations of third party programs typically used installation verification, other field-collected information and free-ridership adjustments. Statewide IOU programs employed installation verification, other field-collected information, free-ridership adjustments and spillover adjustments. However, with the exception of the use of installation verification approaches, less than half of the program kWh savings were evaluated with these adjustment methods.

Figure 14. Energy Savings Evaluation Methods Used by Program Type (kWh Savings)



Verification of Installation and Consumption Analysis

Accurately confirming deemed savings estimates requires verification of installation of program measures and assessment of the actual savings achieved by them. Installation/adoption verification may be obtained directly through on-site visits and telephone surveys or indirectly through consumption analysis showing reduced usage. However, the EEPM directs evaluators to "reference the appropriate IPMVP option" used in their study and to "state any deviations from [the] IPMVP approach."[8] The IPMVP requires on-site verification (and measurement) and does not allow for verification efforts conducted via telephone surveys or interviews.[9] Despite the

[8] Energy Division, 31.
[9] See Chapter 7 of the *Evaluation Framework* for a discussion of measurement and verification and the role and content of the IPMVP.

requirement to use on-site verification, several evaluations relied on telephone surveys and interviews to verify installation, which are not sufficiently rigorous approaches.

A majority of the kWh energy savings goal (55 percent) were evaluated using sufficiently rigorous methods, including both on-site verification and verification through consumption analysis. 4 percent of the kWh savings goals were not evaluated using verification of installation or use consumption analysis (which implicitly includes installation verification) (Figure 15). 23 percent of natural gas savings (Figure 16) and 13 percent of demand savings were evaluated using these methods (Figure 17).

Figure 15. Rigor of Methods Used to Evaluate kWh Impacts (by kWh Savings Goal)

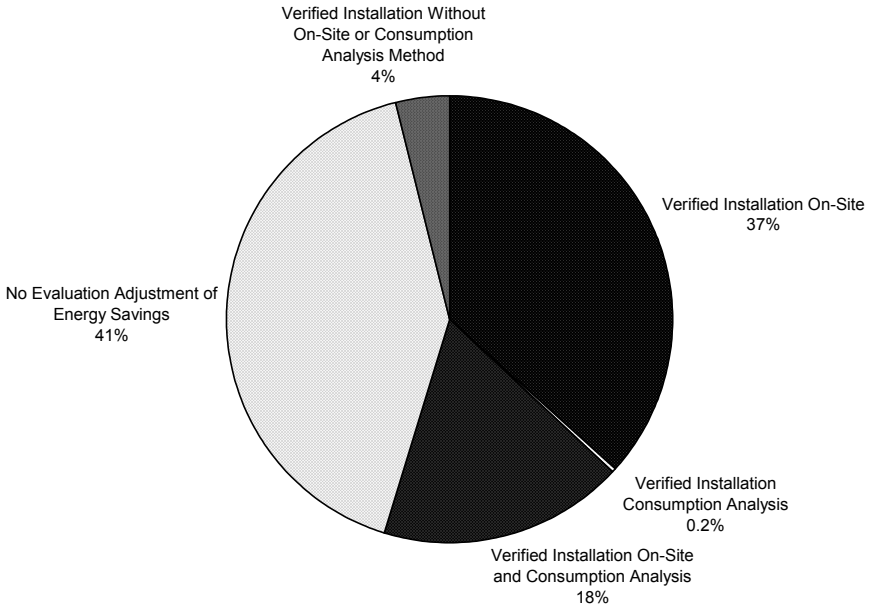


Figure 16. Rigor of Methods Used to Evaluate Therm Impacts (by Therm Savings Goal)

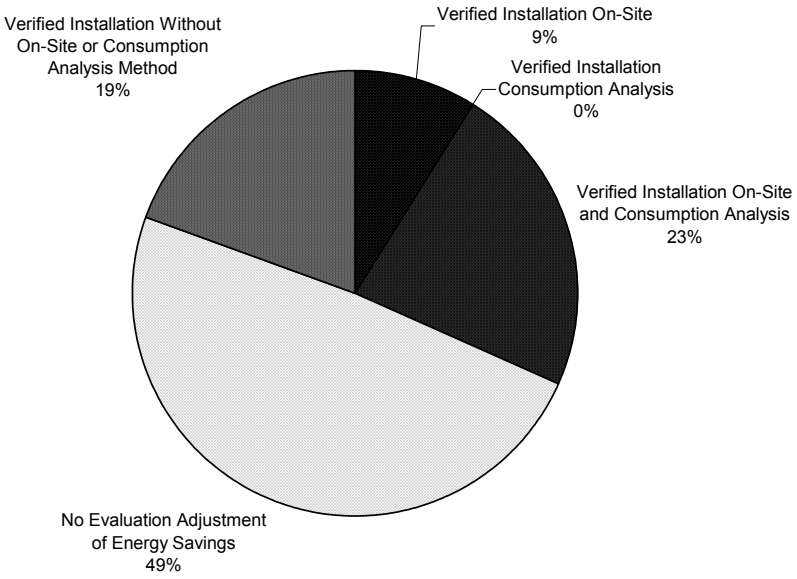
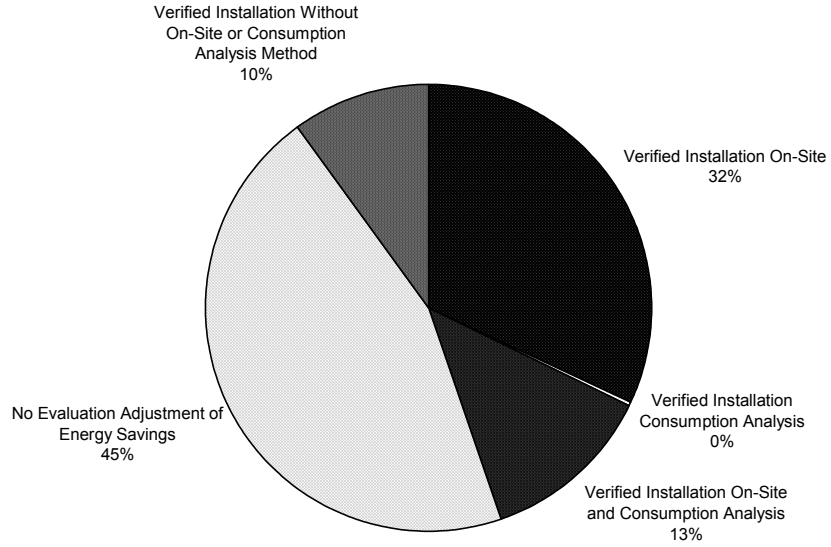


Figure 17. Rigor of Methods Used to Evaluate kW Impacts (by kW Savings Goal)



Net-to-gross Analysis

All of these data collection and adjustment methods should have been directed at developing an estimate of achieved energy savings and comparing it to what would have occurred without the program. As noted in the *Evaluation Framework*:

> In all of these cases what is really desired is the energy and demand savings induced by the program. In other words, the savings need to be "net" of what would have occurred in the absence of the program. Hence, we need to answer the question of what would participants (and non-participants) have done in the absence of the program that could affect their energy use level.[10]

The *Evaluation Framework* also points out that accurate and unbiased estimates have been difficult to acquire through the many methods tested. However, it also highlights the importance of obtaining net savings from the impact evaluation effort. Evaluations that do not provide net impacts are of little value for determining the actually resources available from the portfolio. The *Evaluation Framework* net-to-gross discussion notes:

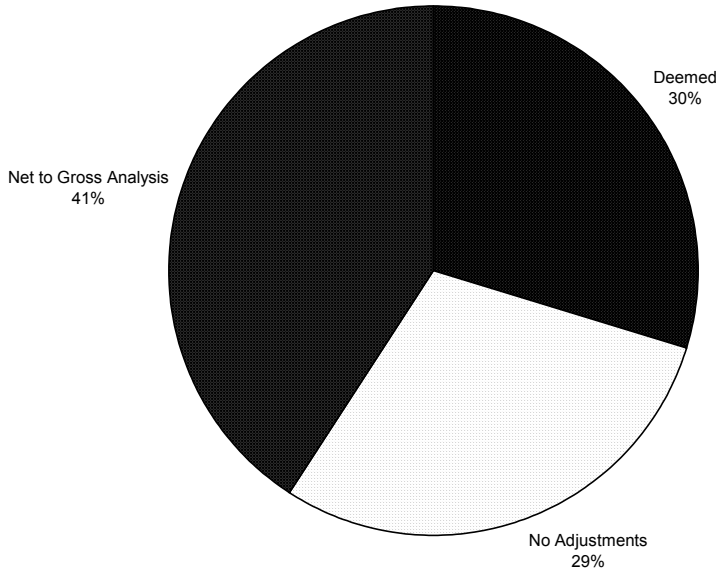
> There are many econometric methods and survey methods that can be applied to the question of estimated NTGR [net-to-gross ratio]. Most, if not all, contain issues with potential bias, limiting assumptions, or lack of outside verification and testing. On the other hand, many methods may provide defensible estimates for a wide variety of specific programs. The Framework recommends that a method be selected in the evaluation planning process that is consistent with the type of program, its program theory, knowledge of the data collection effort that can be accomplished, and a critical approach to measurement and assessing potential problems.[11]

The majority of the reviewed 2002-2003 Portfolio evaluations conducted a net-to-gross analysis that was at least informed by survey information collected from program participants. However, a large portion of the 2002-2003 Portfolio efforts did not have net-to-gross adjustments applied to energy savings.[12] Seven resource acquisition programs did not make a net-to-gross energy savings adjustment (29 percent of the resource acquisition 2002-2003 Portfolio expenditures), representing 42 percent of kWh savings, 50 percent of therm savings, and 46 percent of demand savings.

Specific net-to-gross values were not always clearly presented in the remainder of the reports. In certain cases, deemed estimates were used and in others the evaluators applied an appropriate net-to-gross adjustment based on their research findings. In order to perform a net-to-gross analysis the evaluators had to at least collect some form of participant-provided data to estimate free-ridership levels. Figure 18 presents the percentage of resource acquisition program expenditures with impact evaluations including some form of net-to-gross analysis. 30 percent of program expenditures had evaluations that used a pre-determined "deemed" net-to-gross adjustment value. 41 percent of the program expenditures had evaluations that used at least survey data to develop and apply a net-to-gross value, and 29 percent of the program expenditures were evaluated without any net-to-gross adjustments to the reported savings.

[10] TecMarket Works, 133.
[11] Ibid, 145-146.
[12] The adjustment criteria include verification of installation, collection of field information, free-ridership, participant spillover, non-participant spillover, market effects spillover and education program contributions.

Figure 18. Method to Adjust Energy Savings (by 2002-2003 Portfolio Expenditure)



Free-ridership

Program savings can also be influenced by outside factors such as free-ridership. Measuring free-ridership means taking into account those individuals who were going to implement a program measure without the program, but who participate in the program anyway. Their participation and resulting savings should not be credited to program achievements.

Less than half (23 of 50) of the program evaluations took free-ridership into consideration, covering 29 percent of kWh savings, 28 percent of demand savings and 45 percent of natural gas savings goals. The exact degree to which these omissions lead to an over-estimate of energy savings cannot be determined in this review, but it is safe to say that free-ridership may have had an influence in a majority of the programs. However, it should be noted that the issues of identifying free-riders are complicated and estimating highly reliable program-specific free-ridership is problematic at best. This is especially true in states like California that have had strong and on-going programs that may have caused the participant to seek the action taken one or more years before the enrollment date. In many free-rider surveys these "delayed participants" are counted as free-riders because they answer an evaluation question indicating that they would have taken the action without the current program (because past programs had already convinced them to take the action). This is also true for customers who are educated via California's information, education and marketing efforts, subsequently enroll in a resource acquisition program and then answer an evaluation question indicating that the education program was effective because they would have taken the same action without the resource acquisition program.

Positive Impacts on Energy Savings

Several other factors can influence energy savings estimates. Factors such as participant and non-participant spillover can be considered for their positive influence on the adoption of energy-efficient measures or practices. Measuring participant spillover estimates the energy savings from participants who take additional actions because of their positive experience with the program. Non-participant spillover accounts for savings of individuals who did not participate in the program but who were directly or indirectly influenced by the program and, as a result, took measures to save energy. Changes to the way markets operate that are caused by - but not targeted by - the program can also induce impacts. Similarly, the influence of education and information-only programs may affect savings for related resource acquisition programs. These all work to increase the net savings achieved by programs.

Very few of the evaluations examined herein assess such positive net-to-gross factors. Only one included savings associated with market effects-induced impacts. Three evaluations measured participant spillover, three accounted for non-participant spillover, and one took education and information influence into account. When these positive contributions are not included in the assessment and the negative free-ridership effects are, savings estimates can be substantially conservative. This is a likely result for larger programs and programs involving significant education, advertising or training components. Considering that most evaluations did not include the negative adjustments for free-ridership or the positive adjustments for spillover or market effect, the impacts of these two factors may offset each other to some degree.

Precision

Although estimate precision, confidence intervals and other related statistics are important for understanding the reliability of the energy impacts, most evaluation did not provide information on them. Knowing the precision of savings estimates is important for program decision-makers, portfolio planners and policy mangers. Considering that California has selected energy efficiency as an important resource for the state's energy supply, it is critical to know the total supply of energy efficiency resources and to understand the precisions of those estimates.[13]

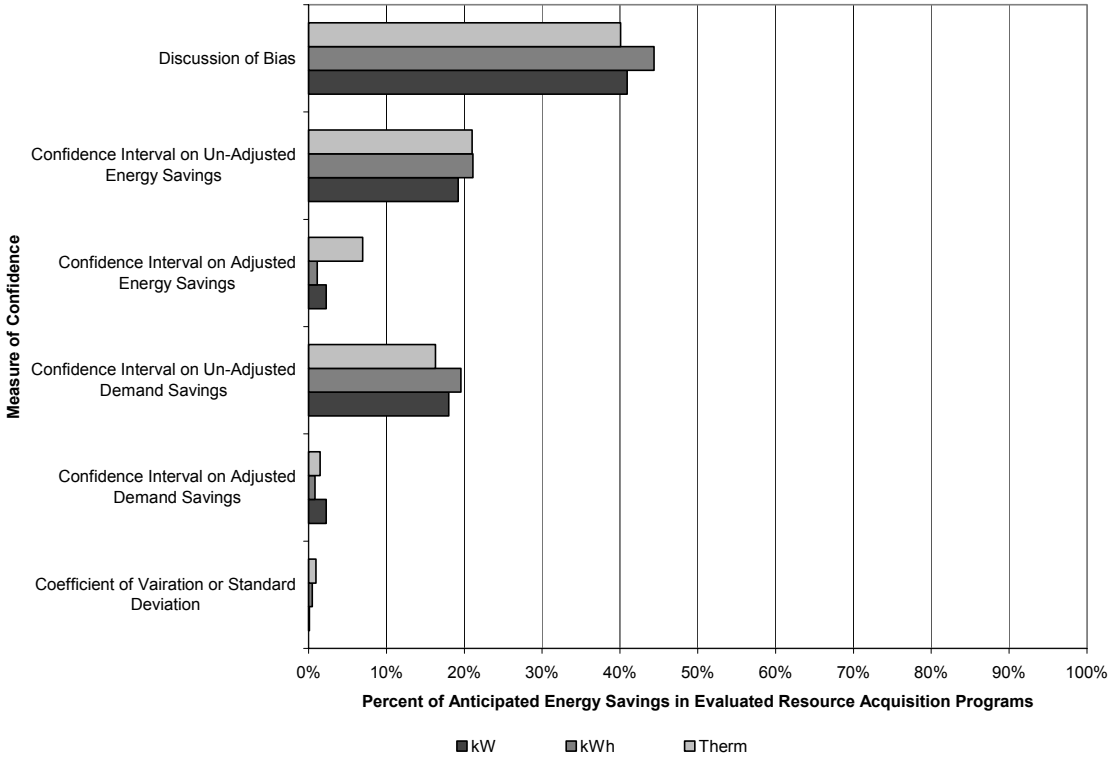
In order to create summary totals along with their precision and error bound levels, program evaluations must provide this information. It would also be helpful for the evaluation reports to present statistics on estimate means, standard deviations and information on the distribution of savings across program components or technology groupings.

Overall, the evaluations of the 2002-2003 Portfolio programs did not sufficiently address issues of precision. Of the anticipated energy savings that were evaluated, 44 percent of kWh and 40 percent of natural gas savings were represented by evaluations that included a discussion of the potential bias in the estimates. Figure 19 presents the distribution of other accuracy metrics contained in the evaluation reports. In cases where bias was considered, discussions were not thorough.

[13] The *Evaluation Framework* provides a discussion, methodology and an example of how program evaluation estimates can be totaled with relative precision and how to calculate the error bound around the total estimate (pages 298 – 300).

Confidence intervals on adjusted net energy savings and adjusted demand savings were only provided on about 1 percent of the savings goal. As discussed above, without these statistics at the program level, precision statistics at the portfolio level cannot be calculated.

Figure 19. Accuracy Assessments within Impact Evaluations

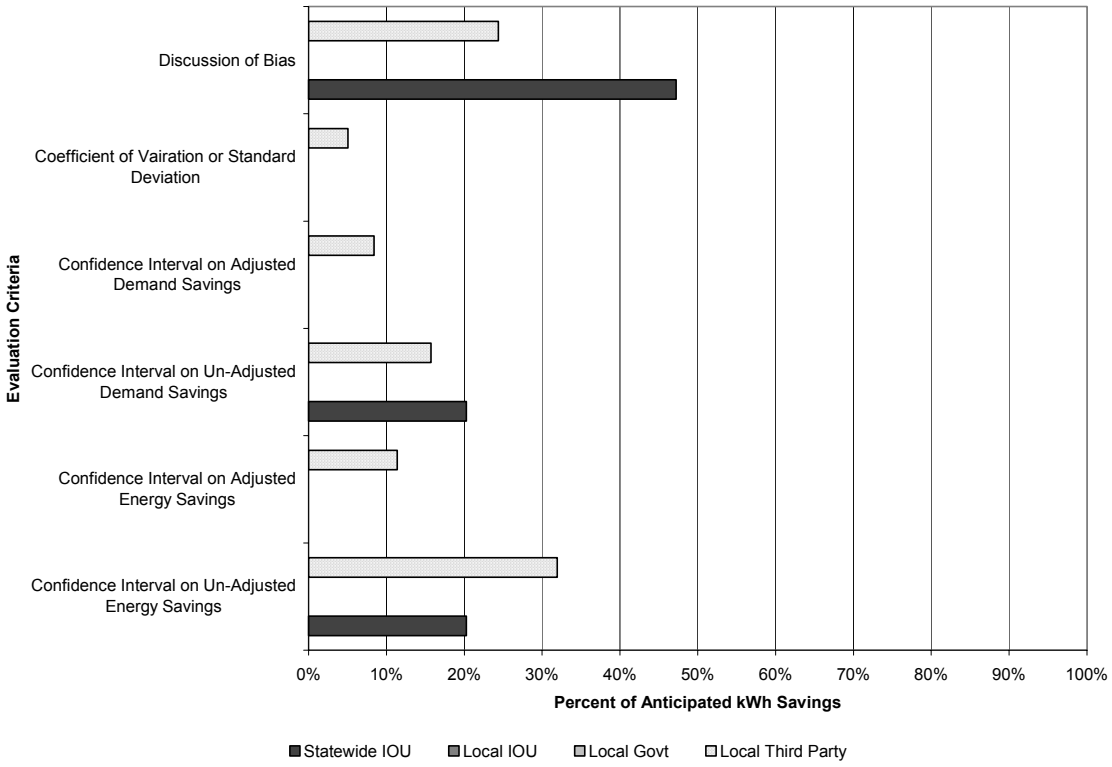


Local third party and statewide IOU program evaluations were the only studies that included any discussion of the confidence level of the impact estimate. The bars in Figure 20 represent the percent of anticipated program savings for each general implementation type. For example, 47 percent of the anticipated energy savings from statewide IOU programs had evaluations that included a discussion of potential issues for bias in the estimates.

Deriving the most efficient sample design for impact evaluations is best done using the estimate reliability metrics from previous studies. When these statistics are not provided it is impossible to know which studies need additional impact assessments and which do not. Evaluation reports need to report these key reliability metrics, including at least some combination of the coefficient of variation, estimate means, standard deviations and/or error ratios for the various sample stratifications used. However, only a small percentage (only 2 of the 50 evaluation reports) of impact evaluations provided these statistics. Even the larger well-funded evaluations did not include these statistics.[14]

[14] See Chapter 13 of the *Evaluation Framework* and its references for recommendations.

Figure 20. Confidence in Impact Evaluation by Implementation Type



Cost-effectiveness

Cost-effectiveness assessments provide valuable information for future portfolio and program planning. Cost-effectiveness is estimated by the IOUs in their annual program reports and program implementation plans. However, it should also be a component of the "arm's-length evaluation" effort because program-projected or program-reported cost-effectiveness may change when evaluation confirmed savings are used. Additionally, more in-depth analysis of the program TRC ratio under alternative program designs or implementation approaches developed within the evaluation study can provide additional information for guiding future program design.

36 percent (18 of 50) of the evaluations for the 2002-2003 Portfolio included evaluation-based cost-effectiveness tests that incorporated impact evaluation results. These reports represent 8 percent of the 2002-2003 Portfolio expenditures for resource acquisition programs (Figure 21). Note in Table 12 that the only studies with cost-effectiveness included in the evaluation reports were the local third-party programs and the local government programs. Though only a small percentage of the impact evaluations provided cost-effectiveness analyses, the EEPM, which the CPUC instructed all evaluators to follow, states that measuring cost-effectiveness is one of the evaluation efforts' primary objectives (page 31).

Figure 21. Percent of Anticipated Energy Savings With TRC Analysis Included in the Impact Evaluation Report

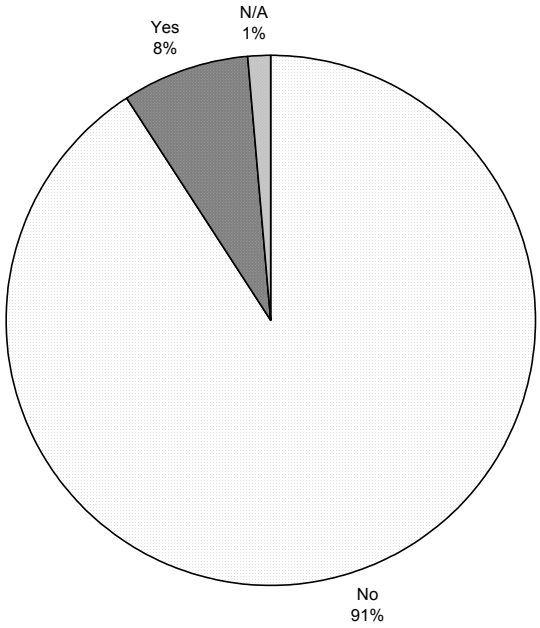


Table 12. Cost-effectiveness Test Included in Impact Evaluations

Program Implementation Type	Percent of Impact Evaluations Including Cost-Effectiveness Test	Percent of Program Expenditures Including Cost-Effectiveness Test
Local-GOV	33%	60%
Local-IOU	0%	0%
Local-TPI	49%	45%
SW-IOU	0%	0%
All Resource Acquisition Programs	36%	8%

Information-only Programs with Impact Evaluation

While impact evaluations were not required for information-only programs, four of the 2002-2003 studies of these programs included such a component in their evaluation research designs. These programs are presented in Table 13.

Table 13. Information-Only Programs with Impact Evaluation

Program Type	Program Name	Evaluated Annual kWh Savings	Evaluated Annual Therm Savings	Evaluated kW Savings
Local-IOU	Residential In-Home Energy Survey Program	2,398,807	0	521
SW-IOU	Codes & Standards	163,490,000	3,935,646	76,630
SW-IOU	Home Energy Efficiency Surveys	24,303,850	1,323,794	0
SW-IOU	Nonresidential Energy Audit	56,147,524	477,192	9,448
	Grand Total	**246,340,181**	**5,736,632**	**86,599**

The Residential In-Home Energy Surveys program evaluation verified actions taken by participants and used a 1995 study's billing analysis and adoption ratios to develop the program savings estimate. This evaluation did not include any discussion of the reliability or confidence of the estimates, but provided a reasonable way to build on past evaluations to estimate impacts of the 2002-2003 Portfolio programs.

The Codes and Standards program evaluation of energy savings first identified the code change proposals developed by the program and successfully adopted into code. Estimates of energy savings from Ely and Associates' *Impact Analysis for the 2005 Update to the California Energy Efficiency Standards for Residential and Nonresidential Buildings*[15] were then used to quantify the 2002-2003 program impacts. This 2005 study thoroughly discussed bias in its approach.

The Home Energy Efficiency Surveys program evaluation looked critically at previous evaluation calculations for the online audit tools. Estimates made for the audit tools in an earlier report were re-assessed for specific end-use technologies and assessed for a typical home's usage to confirm if the previous estimates were realistic. These re-assessed values were then applied to the 2002-2003 programs. In this case, confidence intervals were provided for the estimates and potential estimate bias was discussed in relatively thorough detail.

Gross energy savings impacts for the Nonresidential Energy Audit were based on field information gathered through a participant survey. The results of this effort provided "self-reported" energy savings and only included lighting and cooling measures.

Process Evaluation Efforts

Process evaluation is an important component of the overall evaluation of program performance. It can provide a measure of the effectiveness of the implementation of a program and constructive recommendations about ways to improve program structure, function and operations. As noted in the *Evaluation Framework*:

[15] *2005 Energy Efficiency Standards for Residential and Nonresidential Buildings,*California Energy Commission, Workshop Draft, November 5, 2002, P400-03-001D, Contract 400-00-061. This document can be downloaded from: http://www.energy.ca.gov/title24/2005standards/archive/documents/2002-11-05_workshop/2002-10-22_RES_NONRES_STAND.PDF

> [The] process evaluation consists of in-depth examinations of the design, delivery, and operations of energy programs in order to improve the ability of the program to achieve energy savings and accomplish other program goals.[16]

Process evaluation can be characterized in two dimensions: documenting program operations, and identifying and providing recommendations to improve program efficiency and effectiveness. Evaluation must be detailed enough to provide concrete, actionable material to the program implementer. Again in the *Evaluation Framework:*

> The process evaluation must be conducted at a sufficiently detailed level of investigation to be able to make clear and specific recommendations pertaining to what aspects of the program's management, structure, function, and operations need to be changed.[17]

All but two of the evaluation reports for 2002-2003 Portfolio programs included some form of a process evaluation. In them, implementers had considerable flexibility in identifying the issues on which the process evaluations would focus and on determining the evaluation approach. Most studies, however, were constrained by the percent of the evaluation budget that could be directed to any single process evaluation study. While a wide range of methods were used to evaluate the processes of 2002-2003 Portfolio programs, few of the evaluations were performed at a level of detail supporting the development of detailed recommendations to improve the program operations.

This section of the report first reviews the specific process evaluation issues addressed in the evaluation reports and examines the differences in the issues addressed for the two general types of programs (information-only and resource acquisition) and for each implementation type. Then the overall value of the process evaluations conducted is discussed by examining the recommendations generated and the breadth of the subjects examined. Finally, one process evaluation area of investigation (participant satisfaction) is reviewed more closely to present an example of the level of detail provided within the process evaluation efforts.

Process Evaluation Issues

The process evaluations of the 2002-2003 programs covered a wide range of topics. Both information-only and resource acquisition programs made efforts to evaluate program processes and recommend improvements.

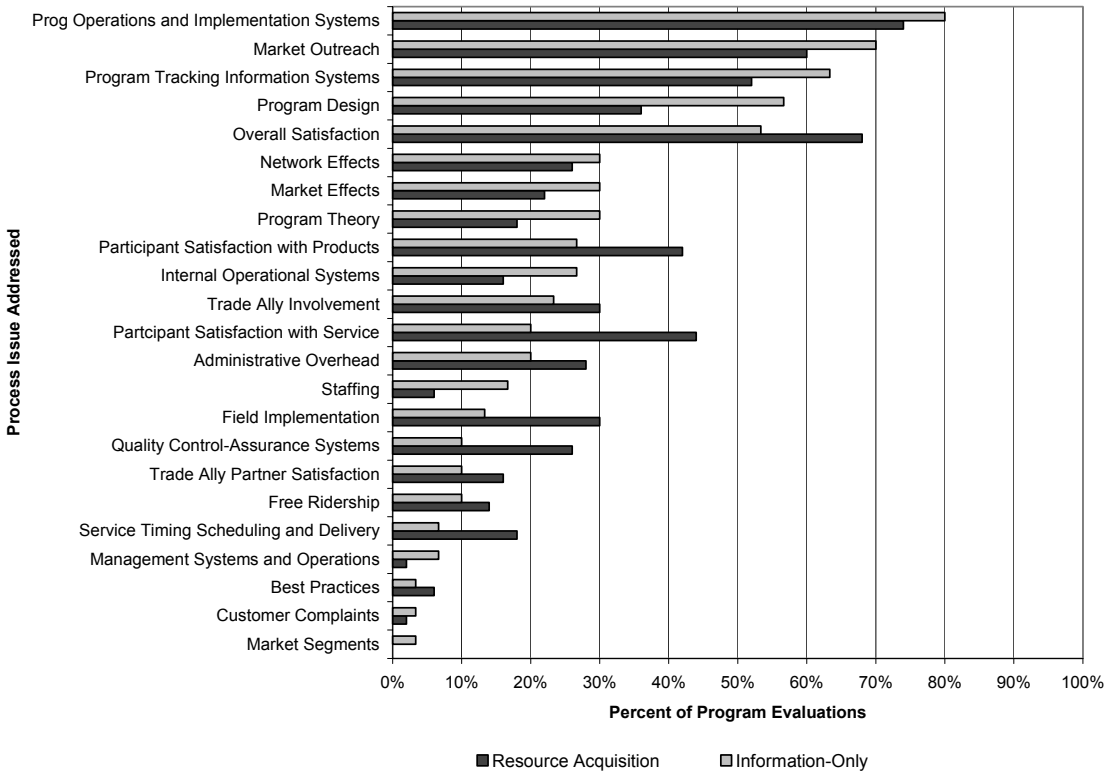
The first series of tables in this section provide an overview of the process evaluation issues covered in the reports. These tables present the issues addressed segregated by general program type (information-only versus resource acquisition) (Figure 22) and by program implementation type (Figure 23 through Figure 26).

Generally speaking, the process issues examined for the information-only programs were similar to those addressed for the resource acquisition programs. This is to be expected as these two types of programs have similar implementation, participation and participant-associated issues. The five most frequently assessed process issues for resource acquisition programs were

[16] TecMarket Works, 205.
[17] Ibid, 212.

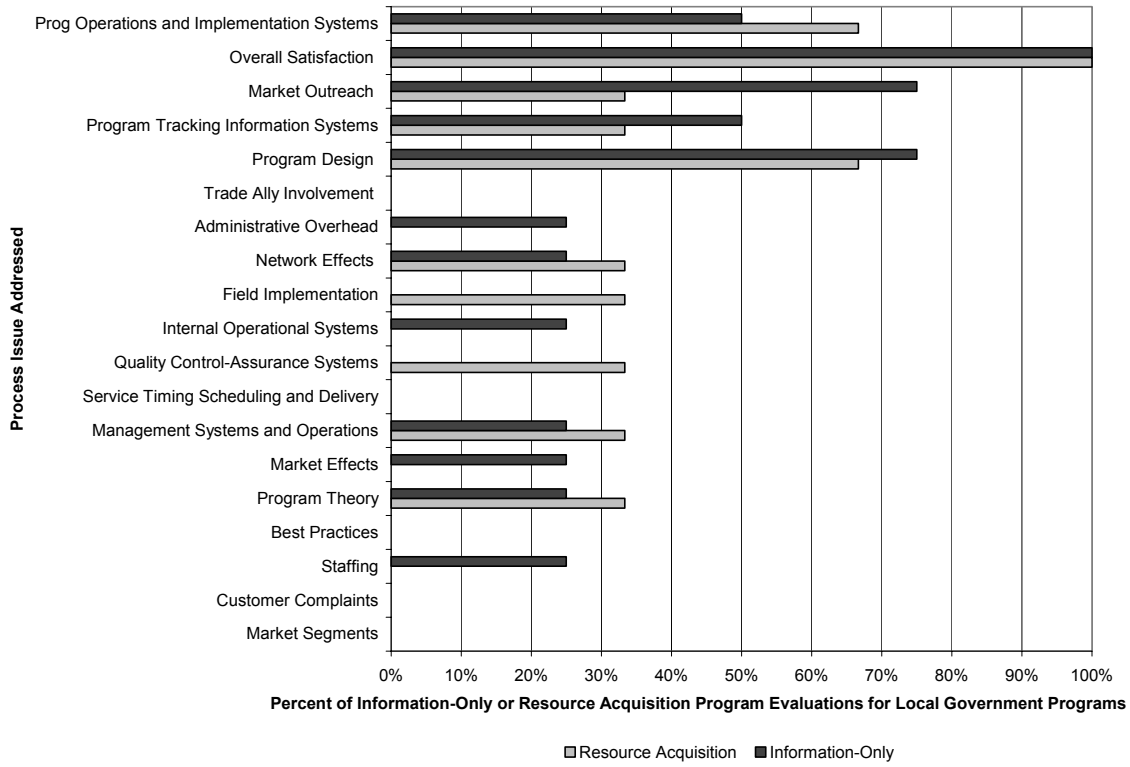
Participant Satisfaction with Service, Program Tracking and Information Systems, Market Outreach, Overall Program Satisfaction, and Program Operations and Implementation Systems. Most of the information-only programs addressed the issues of Overall Satisfaction, Program Design, Program Tracking and Information Systems, Market Outreach, and Program Operations and Implementation Systems.

Figure 22. Evaluation Reports Addressing Process Evaluation Issues



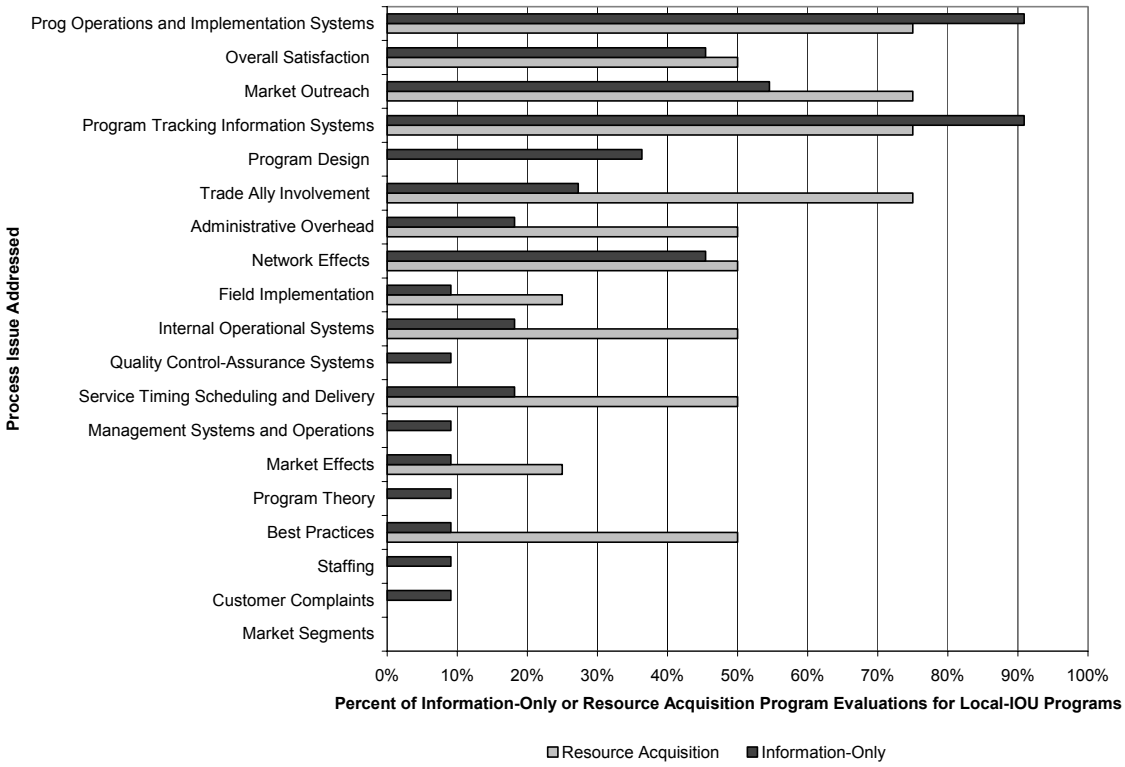
Similarly, there was not a wide variation in the types of process evaluation issues covered for information-only and resource acquisition programs across program implementation types. Again, a significant difference in the issues examined is not to be expected.

Figure 23. Process Evaluation Issues – Local Government



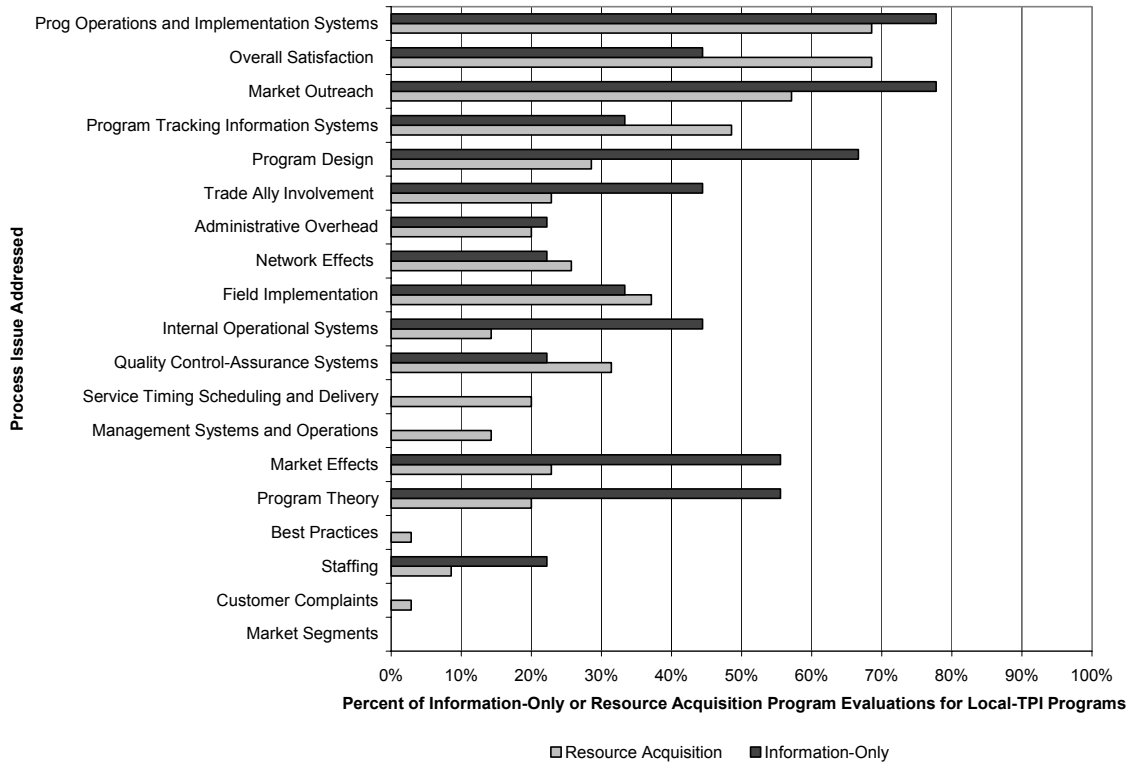
All of the local government program evaluations measured overall program satisfaction, but as will be discussed in the following section the depth of this analysis may not have been sufficient. Program operations, market outreach, program tracking and information systems, and program design were also very commonly included in the process evaluations of the local government programs.

Figure 24. Process Evaluation Issues – Local IOU



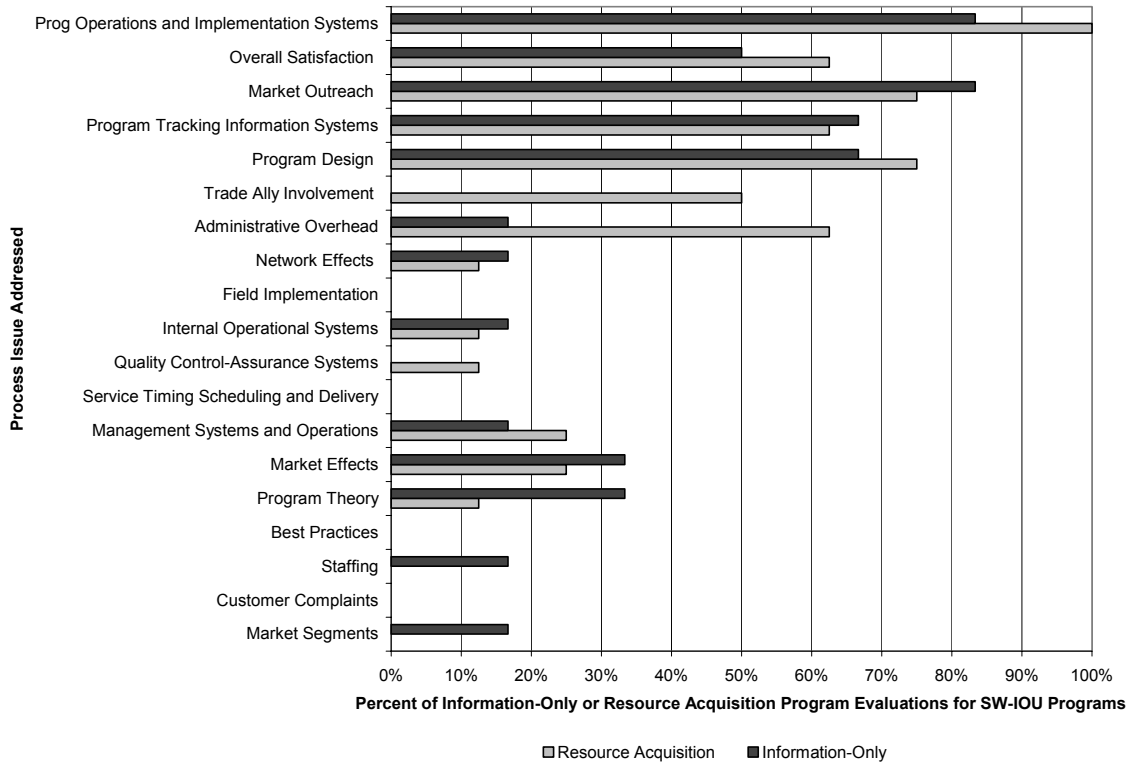
Program operations, implementation systems, program tracking systems, market outreach and overall satisfaction were most commonly evaluated for local IOU programs.

Figure 25. Process Evaluation Issues – Local Third Party



Local third party programs were also evaluated most commonly on the basis of program operations, satisfaction and market outreach. In addition, program design, market effects and program theory also were commonly evaluated components of the information-only programs.

Figure 26. Process Evaluation Issues – Statewide IOU



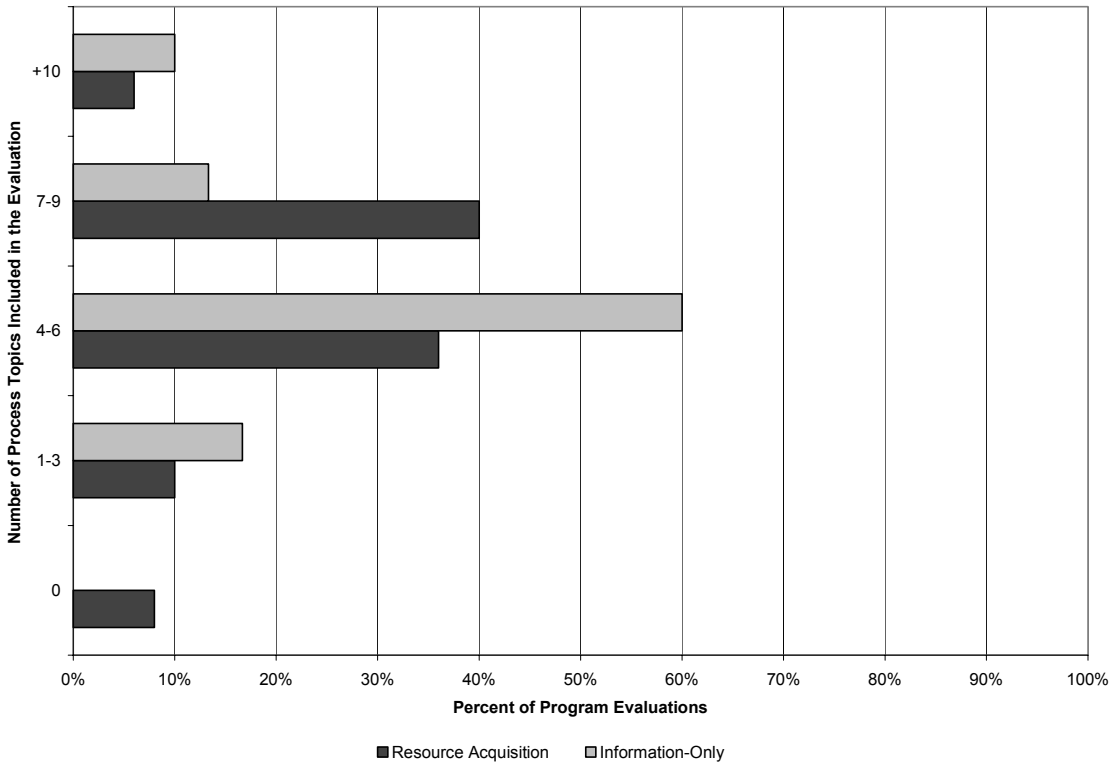
All of the resource acquisition and most of the information-only statewide IOU programs were evaluated on the basis of program operations and implementation systems. Many were also evaluated with respect to market outreach, tracking systems and program design.

Breadth of Process Evaluation

The number of process evaluation issues discussed in its report provides some indication of the breadth of the evaluation effort. As indicated earlier, some of the programs did not conduct process evaluations, while others conducted very limited investigations involving only one or two topic areas. A few others were more comprehensive, covering over 10 different topic areas. In general and as expected, program evaluations with larger budgets tended to investigate more topics than those with smaller ones.

Figure 27 presents the breadth of process evaluations by the number of topic areas examined. Information-only program evaluations typically included 4 to 6 process evaluation topic areas, while most resource acquisition program process evaluations included 4 to 9.

Figure 27. Overall Breadth of Process Evaluations



In looking at the breadth of process evaluations for resource acquisition programs by program implementation type, the majority of program evaluations of statewide IOU and local third party programs included 4 to 9 topic areas (Figure 28). Local IOU and local government resource acquisition program evaluations ranged from 0 to 10 or more. The majority of evaluations included 4 to 6 topic areas regardless of the program implementation type for information-only programs Figure 28.

Figure 28. Breadth of Process Evaluation for Resource Acquisition Programs

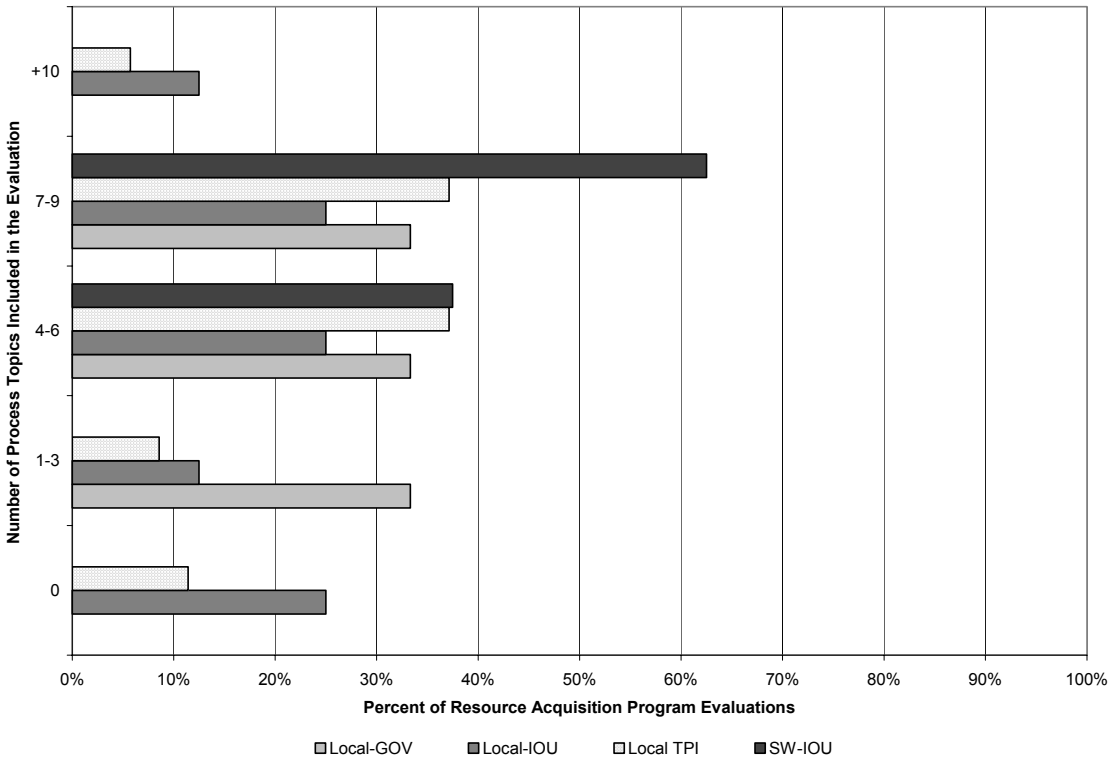
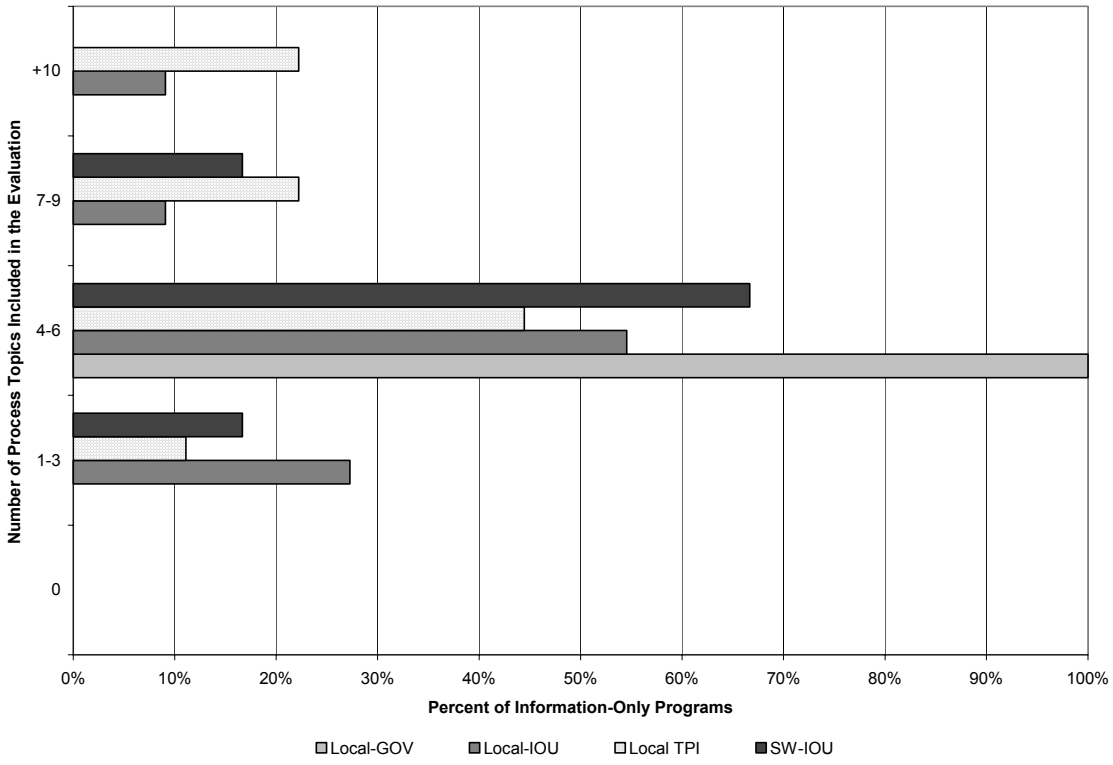


Figure 29. Breadth of Process Evaluation for Information-only Programs



Recommendations Generated by Process Evaluations

The usefulness of any process evaluation depends on its ability to provide an examination in-depth enough to document the program well and identify detailed recommendations to improve program operations or performance. This is especially true for resource acquisition programs, the impacts of which California depends on to meet the state's energy needs. However, information-only programs also need to be evaluated to provide the public with the most cost-effective operations and energy resources possible. Figure 30 and Figure 31 present process issues examined in the 2002-2003 Portfolio program evaluations and indicate if the evaluation provided a recommendation designed to improve program operations or cost-effectiveness.

Unfortunately, few of the process evaluations were conducted in enough detail to fully document program operations. However, several of the evaluations did provide recommendations designed to improve program cost-effectiveness or one or more program components or operational practices. For example, 74 percent of the resource acquisition programs process evaluations examined program operations and implementation systems. 60 percent included recommendations for changes to program operations or implementation efforts. These data indicate that over 80 percent of the process evaluations investigating this area provided recommendations for improvement. However, this also means that 14 percent of the evaluations examined program operations and made no recommendations for improvements, indicating that for these studies in this investigative area the evaluations provided little planning or program change benefits. 68 percent of the studies assessed overall participant satisfaction, yet only 8 percent of the studies provided recommendations to improve it. This is generally due to a lack of detail in satisfaction metrics. This aspect is discussed later in this report.

A considerable percentage of studies provide improvement recommendations for most issues examined in process evaluations. In many ways a successful process evaluation can be defined as one that generates recommendations to improve program operations or performance. From this perspective, most of the process evaluations were successful. However, several topic or issue areas examined had no related recommendations made. In these instances, the process evaluations failed to fulfill their primary objective. However, it should also be noted that there are times when the evaluation contractor finds no room for improvement and therefore makes recommendations. Unfortunately, this is not the case for any of the studies examined in this assessment.

In reading evaluation reports, it is difficult to discern if a lack of recommendations for process issues examined is due to a lack of room for improvement or a lack of detail in the evaluation. Based on a review of the 2002-2003 Portfolio program evaluation plans, the supporting process evaluation budgets and several of the process evaluation field instruments used, it appears that process evaluation results would be significantly improved by assigning specific evaluation budgets to the process evaluation efforts, by hiring firms specifically for their process evaluation skills and experience, and by contractually and managerially separating process and impact evaluation efforts.

Overall, only 11-12 percent of programs on average did not have recommendations to accompany process evaluation issues assessed, This means that on average, recommendations were made for around 90 percent of process issues. The largest gap was in providing

recommendations for overall satisfaction for both resource acquisition and information-only programs.

Figure 30. Resource Acquisition Program Evaluations Making Recommendations for Improvement

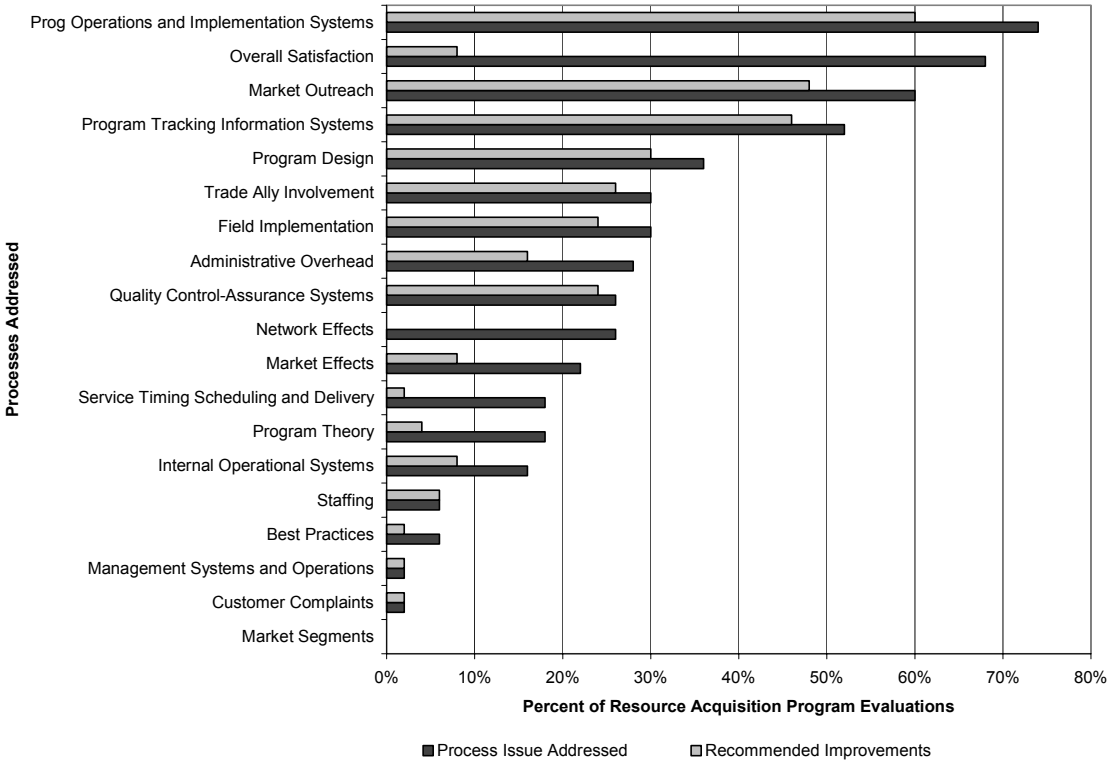
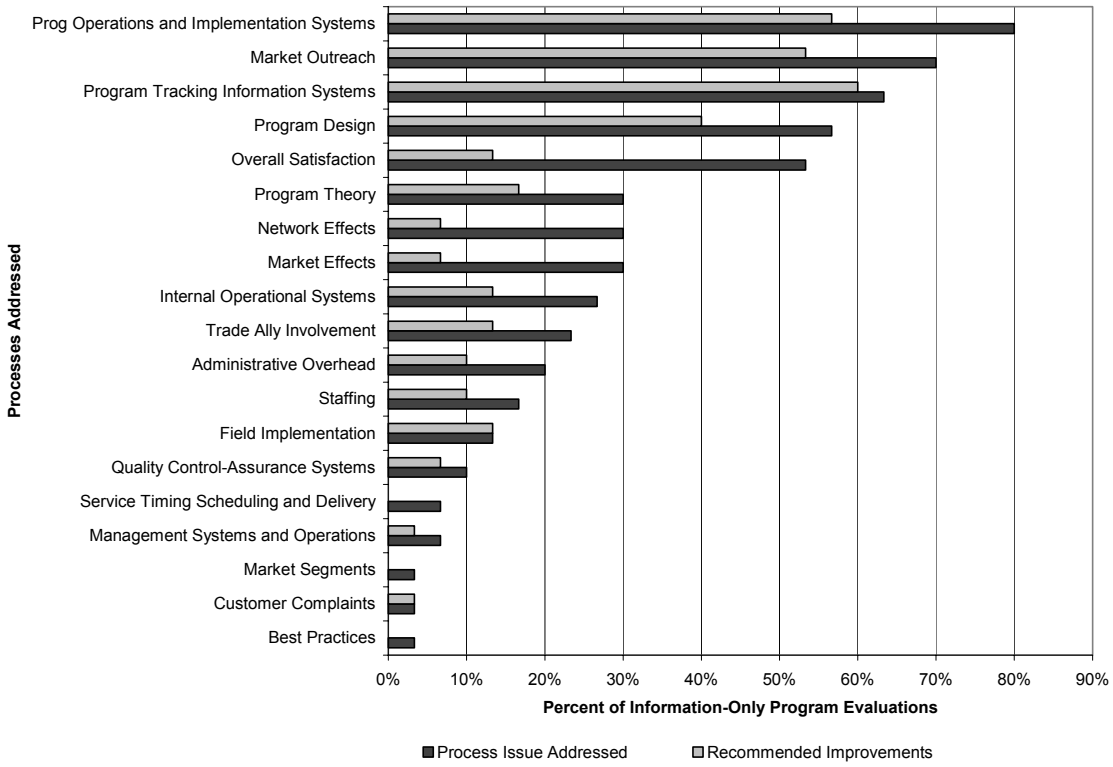


Figure 31. Information-only Program Evaluations Making Recommendations for Improvement



Overall Satisfaction: An Example of Process Evaluation Rigor

One of the most frequently investigated process issues was the measurement of program or program-component satisfaction. However, this was also an area where very few recommendations for improvement were reported. This disparity is due in part to the fact that satisfaction scores are frequently intended for use in ways opposed to those of process evaluations. Program implementers use them as an on-going tracking metric driving "course correction" throughout the program period. Program evaluators, on the other hand, conduct process evaluations only periodically, as they are specifically designed to identify areas where changes are needed and develop recommendations for improvement and are not intended to be a constant program monitoring and management tool.

Only about a quarter of the process evaluations of the 2002-2003 information-only programs generated recommendations for improving satisfaction. A little over 10 percent did so for resource acquisition programs An optimistic assessment of this data is that these programs have reached their satisfaction zenith and only 10 percent of the acquisition programs have room for improving satisfaction. An alternative assessment is that the measurement approaches used were inadequate to provide the detailed information needed to assess satisfaction and develop recommendations. This review of how evaluations treated satisfaction is provided as evidence of the process evaluation rigor employed to measure satisfaction scores.

Certainly, in order to make valid recommendations for improvement in any area, including satisfaction, sufficiently detailed research is needed. For the 2002-2003 Portfolio, 58 of the 80 evaluations (73 percent) include some type of inquiry into participant satisfaction.

In program evaluations including some form of participant satisfaction measurement, "overall" satisfaction questions (i.e., "overall, how satisfied are you with…") were the most common type used.[18] More in-depth analysis of other components of satisfaction, like service, product or service delivery approaches, were less commonly assessed.

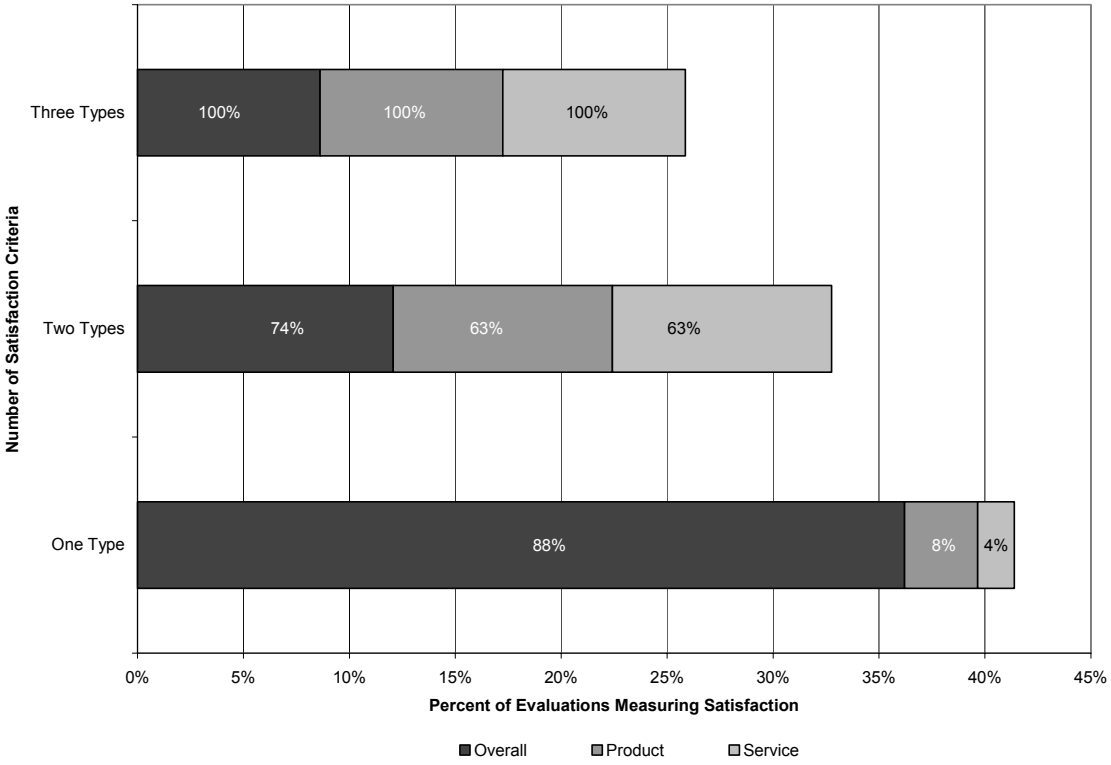
Three types of satisfaction were principally evaluated: overall program satisfaction, satisfaction with the product obtained, and satisfaction with the service provided by the program. These satisfaction questions were not tiered, i.e., an overall satisfaction question was not broken down to questions about the product or service received. A program evaluation may have included a question about product satisfaction without asking about overall satisfaction. Figure 32 presents the percentage of programs evaluating one, two, or three elements of satisfaction in each bar, which is also divided into the various types of satisfaction addressed and the percentage of that type of question.

Represented in the middle bar in Figure 32, 33 percent of evaluations addressing satisfaction asked about two of the three satisfaction components (overall satisfaction, product satisfaction and service satisfaction). Overall satisfaction was the most common, with 74 percent of these studies measuring it in addition to one other type. The top bar represents the 26 percent of studies that asked about all three types of satisfaction.

41 percent of the evaluations addressing satisfaction asked only one type of satisfaction question (bottom bar of Figure 32). Of this 41 percent, 88 percent asked only about overall program satisfaction, 8 percent asked only about satisfaction with the product obtained and 4 percent asked only about service satisfaction.

[18] This did not include a review of the actual satisfaction questions that were asked, but instead how the evaluators characterized their findings. For example, they reported an overall satisfaction level, a participant satisfaction level or a product satisfaction level.

Figure 32. Breadth of Satisfaction Exam Within the Process Evaluation



Often these studies did not examine satisfaction issues with enough rigor to provide recommendations. Effective evaluation of this program component can take significant effort beyond a simple ranking of the general satisfaction responses of high-level program components. A review of the satisfaction questions revealed that many evaluations focused on general satisfaction issues and did not go deeper to focus on issues at the level needed to make program improvements or to identify drivers of low or high satisfaction levels. Similarly, no satisfaction research focused on the three primary relationships that influence satisfaction (customer-product relationships, product-provider relationships and customer-provider relationships).[19] As a result, only a few evaluations provided satisfaction information in enough detail to be of value to program designers or to be useful for understanding what program design or delivery changes could be implemented to improve customer satisfaction. Thorough analysis of participant satisfaction should address design and delivery issues from multiple perspectives and various degrees of detail to arrive at a measure of satisfaction that can be directly used to change program components or operational practices. There is clearly a need for evaluators to design and ask satisfaction questions that can lead to recommendations to improve program performance.

Completeness, Rigor and Evaluation Expenditures

The two preceding sections of this report identify conditions relating to the completeness and rigor of the evaluations of the 2002-2003 Portfolio. This section examines the causes of these

[19]Nick Hall and John Reed, *Methods for Measuring Customer Satisfaction* (Chicago: IEPEC 1997 Conference Proceedings).

conditions. Some of the conditions reported above can be directly attributed to the lack of an evaluation protocol to guide the scope and quality of the research efforts. As discussed earlier, administrators were directed by the CPUC to use the EEPM to guide their evaluations, an activity which it was not designed to support. The EEPM provides little direction on how evaluations should be conducted. Additionally, administrators were allowed to set their own evaluation budgets during the program submission and approval process. In this environment, the evaluation budget was counted as part of the program cost and therefore acted to lower the apparent cost-effectiveness of the programs.

Several administrators indicated that this caused them to submit as low an evaluation budget as possible. They realized that if they were to allocate an evaluation budget to match the research objectives of the EEPM, their proposal would be less competitive than other proposals allocating a minimum evaluation budget. As a result, the 2002-2003 Portfolio evaluation budgets were, in most instances, significantly below the level needed to comply with all the evaluation objectives required by the CPUC. This proposal and budget development process caused many of the 2002-2003 evaluation efforts to employ strategies incapable of providing reliable results. Because program budgets and budget allocations had already been approved by the time detailed evaluation plans were developed, evaluation budgets allocated during the program planning process were typically considered evaluation spending caps. In addition, in many cases a significant portion of the already low budgets allocated to the evaluation efforts were spent by the administrators contracting the evaluation rather than on the evaluation itself.

This section examines if a lack of funding affected evaluation quality.

Compared to the administrator-recommended levels of funding for administering, managing and implementing the evaluation efforts, the 2002-2003 Portfolio programs were highly variable in their total spending on evaluations. The *Evaluation Framework*[20] recommends evaluation spending be 4 to 10 percent of the program budget, depending on the specific circumstances of the programs being evaluated and the evaluation needs of the specific program.

Spending on the 2002-2003 evaluation efforts ranged from 0.04 percent to 24 percent, with an average of about 5 percent (Table 14). A majority (56 percent) of the programs fell into the recommended 4-10 percent evaluation expenditure range. The statewide IOU programs had the lowest proportion of spending on evaluations, but these studies often also had larger evaluation budgets because of the size of the program budget as a whole.

[20] TecMarket Works, 75.

Table 14. Program Spending on Evaluation

		Evaluation Expenditures as a Percent of Total Program Expenditures		
	Type	**Average**	**Max**	**Min**
Information	Local-GOV	4%	4%	3%
	Local-IOU	5%	14%	1%
	Local-TPI	4%	15%	1%
	SW-IOU	4%	6%	2%
Information-only Total		**5%**	**15%**	**1%**
Resource Acquisition	Local-GOV	7%	15%	2%
	Local-IOU	4%	6%	1%
	Local-TPI	5%	24%	0.04%
	SW-IOU	2%	4%	0.34%
Resource Acquisition Total		**5%**	**24%**	**0.04%**
All Programs		**5%**	**24%**	**0.04%**

Table 15. Evaluations Meeting Recommended Evaluation Funding Level

	Type	**Under 4%**	**Between 4-10%**	**Over 10%**
Information	Local-GOV	44%	56%	0%
	Local-IOU	13%	63%	23%
	Local-TPI	27%	37%	36%
	SW-IOU	14%	86%	0%
Information-only Total		**21%**	**59%**	**20%**
Resource Acquisition	Local-GOV	10%	21%	69%
	Local-IOU	25%	75%	0%
	Local-TPI	23%	59%	18%
	SW-IOU	73%	27%	0%
Resource Acquisition Total		**25%**	**53%**	**20%**
All Programs		**23%**	**55%**	**20%**

It should also be noted that the 2002 statewide IOU program evaluations were conducted during a period in which the focus of evaluation efforts had shifted from performing market effects, market operations and process evaluations to conducting impact evaluations without the use of an evaluation standard or protocol. As a result, the related energy impacts presented in this summary report should be considered as reasonable attempts to quantify net ex-post energy impacts within the budgets and evaluation priorities associated with the 2002 programs.

Merely looking at it as a percent of total program expenditure, however, does not always provide a clear indication of the adequacy of the evaluation budget. There are levels of fixed and variable costs associated with any evaluation study. There can also be greater evaluation budget needs for new or pilot programs that have to be tested, for programs that make up large proportions of the energy savings goals or for policy-support reasons, such as when a program approach is being considered as an implementation requirement. At the same time, there are very significant economies of scale that can occur. This is particularly true for sampling when the required sample size is more dependent on the variation in the item of interest than in the size of the participant population. Similarly, the costs of conducting a regression-based consumption

analysis from utility bills is much more dependent on the methods employed than on the sample size used.

While impact evaluations are generally required at least periodically for resource acquisition programs, the costs for a proper impact evaluation can vary significantly by methodology, size of program (particularly the case for large multi-measure programs), and expenditures on measurement and verification (M&V). At the same time, there is a minimum level of expected expenditures if there are to be enough on-site M&V visits to support an impact evaluation.

With regard to costs for process evaluations, the *Evaluation Framework* states:

> [T]ypically process evaluations range in cost from a low of around $10,000 to $15,000 for small evaluations of limited scope, to an average cost of around $30,000 to $60,000 for evaluations focusing on multiple issues with some level of supportive field data collection and detailed assessment efforts. However, it is not unusual for process evaluations that focus on a wider range of issues involving customer surveys, interviews or focus groups with on-site examinations to run in excess of $60,000.[21]

It is highly unlikely, then, that an adequate impact and process evaluation designed to meet EEPM requirements could cost less than $60,000, even for a small program. More likely, a minimum budget level would be $70,000 with a more reasonable estimate being over $80,000, and with major program evaluation budgets being well into six figures.

18 of the 2002-2003 Portfolio programs (21 percent) had evaluation budgets under $20,000. 41 programs (48 percent) budgeted less than $50,000. Most of these evaluations were not sufficiently funded to meet the goals described in the EEPM as required by the CPUC.

Table 16. Distribution of Evaluation Spending by Program Type (Percent of Programs)

Type	Evaluation Spending Range				
	Under $20k	$20k-$50k	$50k-$100k	$100k-$200k	Over $200k
Local-GOV	57%	0%	43%	0%	0%
Local-IOU	38%	38%	25%	0%	0%
Local-TPI	15%	35%	31%	17%	2%
SW-IOU	0%	14%	21%	21%	43%
Grand Total	**20%**	**29%**	**29%**	**13%**	**8%**

A better way to assess the adequacy of the evaluation budgets is to compare evaluation spending to evaluation quality. In Figure 33 evaluation expenditures are compared to evaluation rigor. Level of rigor, it should be remembered, typically drives result reliability. There is a strong correlation between expenditure and level of rigor provided for impact evaluations. There were, however, minimally funded but very rigorous studies that complied with IPMVP field data metering and monitoring measurement requirements. On the other hand, there were also expensive evaluations, costing well into six figures that did not comply with the IPMVP, conducting no field metering or other IPMVP-approved impact adjustment approaches.

[21] Ibid, 228.

However, on average, across all studies conducted, the evaluation budget correlates well with evaluation quality.

This correlation is not clear for the process evaluations examined herein (Figure 34). The rigor metric used here is the number of issues examined in the process evaluation. However, more important to assessing value in this case is quality of issue examination rather than the sheer number of issues addressed. A process evaluation can cover one or two issues well and be valuable or cover ten issues poorly and be of no value. This process graphic is presented to demonstrate that there is little correlation between evaluation budgets and number of process issues examined.

Figure 33. Evaluation Expenditure versus Verification Rigor for Resource Acquisition Program Evaluations

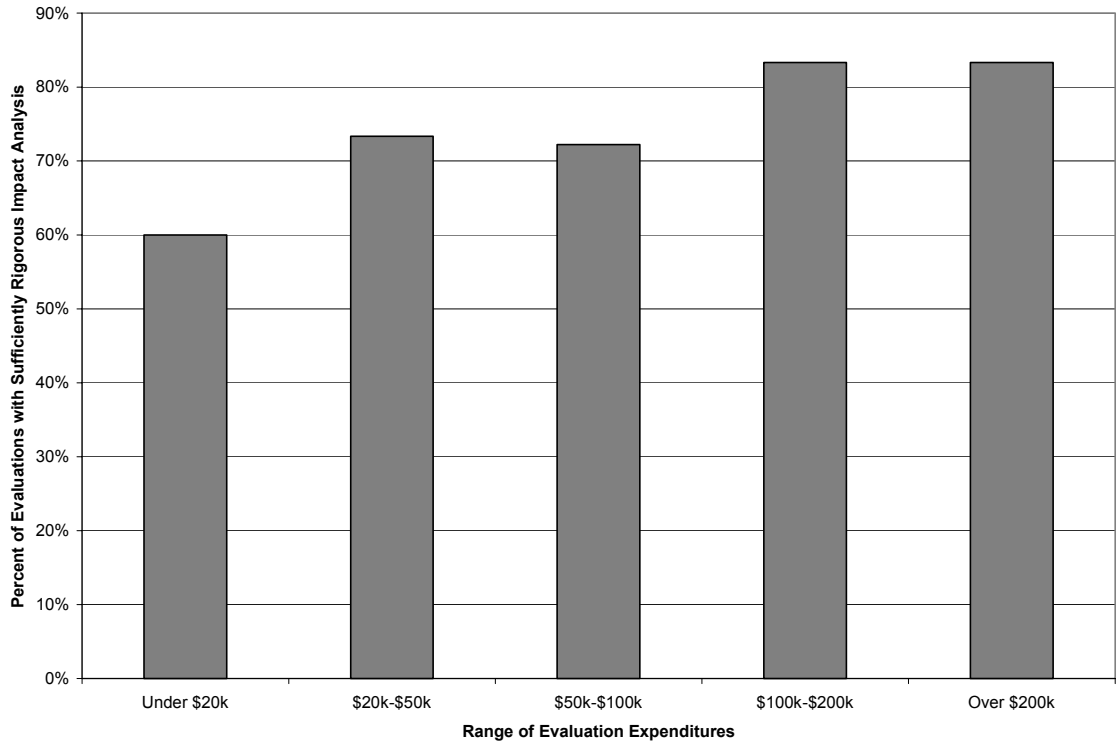
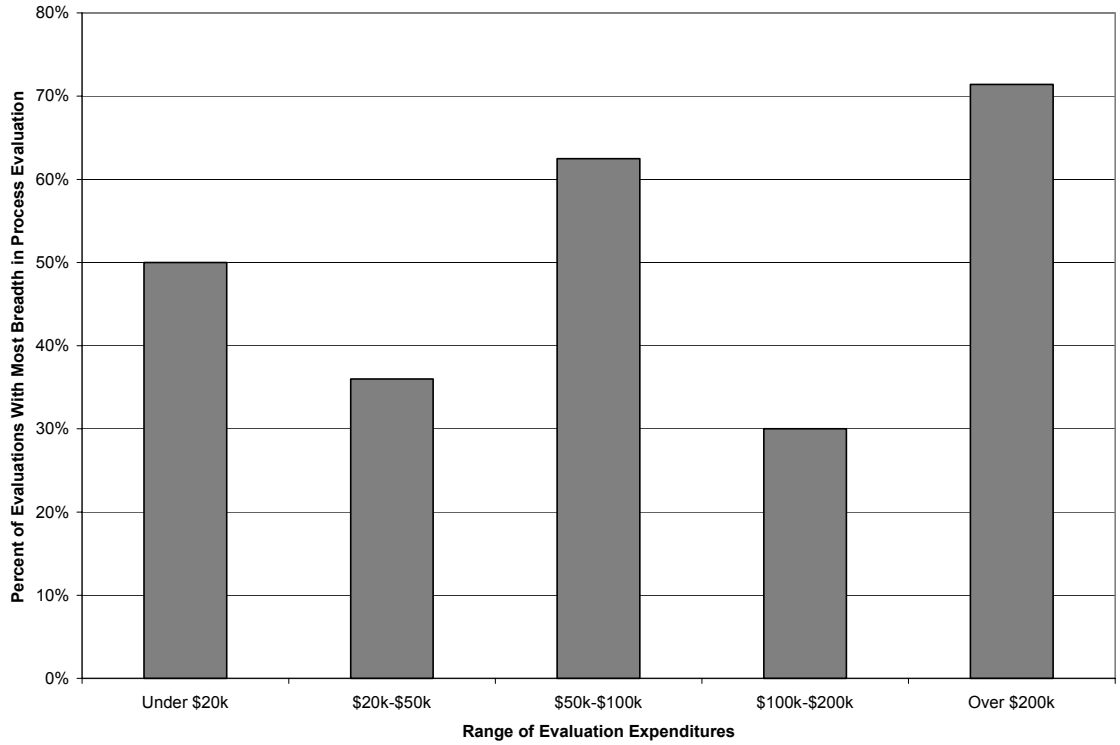


Figure 34. Evaluation Expenditure versus Process Evaluation Completeness



Setting percentage-based minimum levels of evaluation expenditure to meet minimum evaluation expectations can present problems for the smaller programs in a portfolio of offerings that vary considerably in size. To help ease this problem by lowering evaluation costs, evaluations can be conducted periodically or smaller programs can be evaluated with other similar programs. However, cost savings are minimal if results must be reported for each program. A minimum level of consolidation was conducted for the 2002-2003 evaluations and is also occurring for the 2004-2005 studies. For 2002-2003 there were four consolidated program evaluations covering eight programs. While consolidating evaluations may help, it seems apparent that if reliable information is to be provided from the evaluation effort higher percentages of evaluation dollars may be needed for smaller programs. In the long run, focusing on evaluation budgets is likely less important than concentrating on setting minimum levels of study reliability for each program and then structuring evaluations to meet these requirements. Protocols that allow for multiple levels of rigor also provide for multiple levels of reliability. The post-2005 evaluation protocols currently being developed should address this issue..

Desired Comparative Analysis

The original RFP and Work Plan for this project included a significant effort for a summary analysis, comparative analysis and potentially a meta-analysis. Yet, as can be seen in the above discussions, the evaluations were too often not comparable and not rigorous enough to support these types of analyses.

The desired summary study would present a synopsis of energy and demand savings obtained with related confidence levels and discussions of potential sources of bias and overall independent assessments of estimate reliability. In order to do this, impact evaluations would need to have a minimal level of verification and savings analysis and report confidence levels, standard deviations and potential bias assessments. Few of the 2002-2003 evaluations actually met these criteria. This will likely be true for the 2004-2005 evaluations as well, as these elements are not being required of all impact evaluations and the evaluation budgets were set in a fashion similar to the 2002-2003 studies. This need for reported statistical reliability should be considered during evaluation planning for the 2006-2008 California programs.

Proxy Cost-Effectiveness Test

While most 2002-2003 evaluations do not report program cost-effectiveness, the final program workbooks provided to the CPUC do. The cost-effectiveness calculations in these final workbooks are based on actual program expenditures, actual program accomplishments (in terms of measure installations) and ex-ante unit savings estimates. To conduct an evaluation-based ex-post net impact proxy cost-effectiveness analysis, the realization rate from the 2002-2003 impact evaluations was used to adjust the net present value (NPV) of total resource benefits presented in the final workbooks. Since the evaluations of the 2003 IOU programs had not been completed at the time of this report, the realization rates from the 2002 program evaluation studies were applied to the 2003 claimed savings to estimate the 2003 evaluation adjusted savings. These were then summarized by program type. The NPV of benefits and costs was then summarized for the 2002-2003 Portfolio. Since the NPV of the energy savings benefits is a function of the life of the measure, a program average effective useful life (EUL) was calculated from the measure level energy savings and EUL data contained in the final workbooks. The program average EUL was calculated as an energy savings weighted average of the effective useful lives of the measures installed by the program. The evaluation-adjusted NPV of the total resource benefits for the TRC test was calculated from the utility avoided cost data provided in the final workbook at the program average EUL and the evaluation-adjusted annual energy savings.

The evaluation-adjusted NPV of the total resource benefits for the 2002-2003 Portfolio is approximately $1,435,000,000. The approximate total NPV of the total resource costs of these programs is $665,000,000. This provides an approximate total resource net benefit for the full 2002-2003 Portfolio of $770,000,000 and total resource benefit cost ratio of 2.16. The estimated total resource costs and benefits for the 2002-2003 Portfolio are summarized in Table 17.

Table 17 . Proxy Total Resource Costs and Benefits for 2002-2003 Programs

Cost Category	All Programs	Resource Acquisition Programs Only
Total Resource Benefits	$1,435,377,723	$1,435,377,723
Total Resource Costs	$664,752,059	$611,701,940
Total Resource Net Benefits	$770,625,678	$823,675,783
Total Resource Benefit Cost Ratio	2.16	2.35

The cost-effectiveness data for the 2002-2003 Portfolio were disaggregated by program implementation category, customer class, primary market segment served and primary program

delivery strategy. Table 18 presents the proxy measurement of the TRC by implementation category for the 2002-2003 Portfolio.

Table 18. Cost-Effectiveness of 2002-2003 Programs by Implementation Category

Implementation Category	Proxy TRC (All Programs)	Proxy TRC (Resource Acquisition Only)
Local-GOV	0.63	1.28
Local-IOU	0.80	2.04
Local-TPI	1.71	1.91
SW-IOU	2.32	2.42
All Programs	**2.16**	**2.34**

According to this analysis, resource acquisition programs delivered by all categories of implementers were cost effective. Resource acquisition programs offered by the IOUs were the most cost effective, and resource acquisition programs offered by third party implementers were also highly so, with a benefit cost ratio near 2.0. The cost-effectiveness of the full 2002-2003 Portfolio offered across implementation categories varies due to the relative mix of information-only and resource acquisition programs offered within each.

The program cost-effectiveness by customer class is summarized in Table 19. As this shows, programs targeting the commercial sector are approximately 1.9 times more cost effective than those targeting the residential sector. Though this is not surprising, it is generally recognized that there are many cost-effective opportunities in the residential sector, and that California policy makers cannot achieve their energy efficiency goals without obtaining significant savings from it.

Table 19. Cost-Effectiveness of 2002-2003 Programs By Customer Class

Customer Class	Proxy TRC (All Programs)	Proxy TRC (Resource Acquisition Only)
Commercial	2.37	2.46
Residential	1.28	1.29
Crosscutting	3.13	4.51
All Programs	**2.16**	**2.34**

Crosscutting programs showed the highest level of cost-effectiveness. This result was driven primarily by the cost-effectiveness of crosscutting Codes and Standards programs, as described later in this section.

Within each customer class, some programs were designed to target specific market segments, while others had a more general focus. The cost-effectiveness of programs designed to address specific market segments is shown in Table 20.

Table 20. Cost-Effectiveness of 2002-2003 Programs By Target Market Segment

Target Market Segment	Proxy TRC (All Programs)	Proxy TRC (Resource Acquisition Only)
Agricultural	0.57	1.15
Hard-to-reach	0.94	1.42
Industrial	2.12	2.12
Local Government	0.96	1.53
Multi-family	1.81	1.98
Mobile Home	1.25	1.25
Single Family	1.70	1.70
Schools[22]	-	-
Small Commercial	2.06	2.33
University	1.80	1.80
All Programs	**2.16**	**2.34**

Targeted resource acquisition program were cost effective in all target markets served. Resource acquisition programs specifically targeting hard-to-reach customers and market sectors, such as small commercial, multi-family and mobile home, tended to be less cost effective than the overall 2002-2003 Portfolio average due to the added costs associated with reaching these segments. The cost-effectiveness of small commercial and multi-family resource acquisition programs, however, approached the 2002-2003 Portfolio average.

Table 21 presents the TRC according to the primary program delivery strategy as identified by the Master Evaluation Team. Some programs used more than one delivery strategy. Programs that used the promotion of codes and standards as a delivery strategy showed the highest levels of cost-effectiveness in the 2002-2003 Portfolio. While codes and standards programs are typically submitted to the CPUC as information-only programs, impact studies conducted for the IOU programs have reported high levels of savings demonstrating their cost effectiveness. Upstream programs were more cost effective than the 2002-2003 Portfolio average, reflecting lower delivery costs relative to programs targeted to end-users. Direct installation programs also showed good cost-effectiveness, registering a higher TRC than traditional prescriptive and custom rebate programs.

Table 21. Cost-Effectiveness of 2002-2003 Programs By Program Delivery Strategy

Program Strategy	Proxy TRC (All Programs)	Proxy TRC (Resource Acquisition Only)
Audits	1.41	1.75
Codes and Standards	31.27	34.19
Direct Installation	2.31	2.31
Education, Training and Information	0.00	0.52
Financing	0.38	0.38
Rebate – Customized	2.16	2.16
Rebate – Prescriptive	1.85	1.85
Performance Contracting	2.14	2.14
Upstream	2.99	2.99
All Programs	**2.16**	**2.34**

[22] All of the schools programs were information-only.

Conclusions, Lessons Learned and Recommendations

Conclusions

The evaluations of the 2002-2003 Portfolio, and consequently the overall evaluation of the 2002-2003 Portfolio, generally suffer from three key problems: incomparability, incompleteness and a lack of rigor. However, despite the overall weakness of the evaluation efforts, there were several very strong program evaluations that provided reliable information. Unfortunately, these were offset by weak studies incapable of providing reliable findings or being used to improve the estimates in the DEER database. As consolation, the most rigorous studies are generally those for the programs with the largest savings, which means that a high percentage of total claimed savings were accurately and adequately evaluated for the 2002-2003 Portfolio, and represent a relatively accurate picture of the actual 2002-2003 Portfolio's savings. The use of the 2002-2003 Portfolio evaluation reports as a resource planning tool, however, is seriously hindered by the above-listed problems across the 2002-2003 Portfolio evaluations.

Incomparability

The incomparability of evaluations, and therefore of program achievements, makes it nearly impossible to draw reliable or detailed conclusions about the contributions of the 2002-2003 Portfolio to meeting California's energy needs. Definitions, evaluation requirements, necessary rigor and general reporting should be consistent across the portfolio to ensure comparability. This will likely improve in future program years given the recent efforts to clarify evaluation guidelines and develop evaluation protocols, and the addition of more oversight and direction in the evaluation planning process.

For example, given lessons learned from the 2002-2003 studies, the CPUC's internal evaluation, measurement and verification (EM&V) plan review and approval process document for 2004-2005 resource acquisition programs includes language to help ensure that evaluation will provide more thorough energy and demand savings estimates. More specifically, it states:

> To meet the CPUC evaluation requirements, the EM&V Plan must have the evaluation:
>
> a. Provide peak savings.
> b. Provide first year kWh and kW savings along with annual savings for years 1 through x depending upon an assessment of measure lifetimes for this program.
> c. Measure net-to-gross and provide net energy and peak savings.

Incompleteness

The 2002-2003 Portfolio programs were not completely comparable because several evaluation reports were incomplete in their assessment and reporting of multiple evaluation components.

While most of the program evaluation studies report first-year savings, the majority exclude an assessment of the expected lifetime associated with those savings or an estimate of the total program savings over time. Some evaluations made comments on EUL estimates and modified these for revised cost-effectiveness calculations, but these efforts were not sufficient to make projections on the lifetime impact of the 2002-2003 Portfolio. The exclusion of lifetime savings

analysis from the evaluations seriously hampers their usefulness and significantly restricts their ability of the evaluation efforts to inform public policy or energy supply decisions. While estimates of the lifetime impacts are provided independently by the IOUs outside of the evaluation process, these metrics should also be provided via the evaluation efforts.

Cost-effectiveness was also inconsistently evaluated or reported. Only 18 of the 50 (36 percent) resource acquisition program evaluations reported evaluation-based cost-effectiveness (or TRC). These programs represent only 8 percent of the 2002-2003 Portfolio expenditures. Evaluation-based TRCs, along with energy and demand savings, are essential to assessing re-investment and program offering decisions.

To be fair, the EEPM in effect at the time of the 2002-2003 program evaluations could be viewed as ambiguous in the area of impact metric definitions. Its Section 6 on EM&V requires "[m]easuring the level of energy and peak demand savings achieved,"[23] but does not specify whether these are first-year or lifecycle savings. It also calls for an analysis of cost-effectiveness which requires an estimate of lifecycle savings. Yet, the EEPM also provides assumed lifecycle and net-to-gross ratios to be used for the program proposal filing. The EM&V section does not explicitly state that these factors need to be measured and used in the evaluations. The final project workbooks all calculate lifecycle savings, which may have been interpreted as satisfying the EEPM requirement. If an evaluation plan omitting cost-effectiveness analysis was approved by the CPUC, it could reasonably be interpreted that performing cost-effectiveness calculations was not required of the evaluation contractor. Year-by-year reported savings were not required by the EEPM or calculated in the program workbooks, so it is logical that none of the evaluations reported results in this manner.

Lack of Rigor

A surprising lack of rigor was found in the evaluation reports as a whole. A full 29 percent of the resource acquisition 2002-2003 Portfolio did not have their energy savings adjusted from estimated gross savings. While the remaining evaluations employed a wide range of evaluation approaches, some studies reported ex-post impacts by applying per-measure deemed or program-estimated savings to the program's tracking system measure counts. This methodology should not be considered an evaluation effort, but instead an accounting approach for confirming estimated savings projections. Other studies used more rigorous approaches that can be expected to provide more accurate and reliable estimates of energy impacts.

The EEPM prescribes a fairly rigorous evaluation standard in its required use of IPMVP evaluation approaches (which require on-site technology-based measurements). However, many evaluators ignored this requirement and did not conduct on-site or site-specific verification, monitoring or metering activities, citing lack of evaluation funds, even though administrators were instructed to budget for IPMVP-based evaluations. In fact only a slight majority of the kWh energy savings (55 percent) was evaluated with sufficient rigor to provide impact estimates with a basic level of reliability. Only 23 percent of natural gas and 44 percent of demand savings were evaluated using reliable methods.

[23] Energy Division, 31.

Accurate impact evaluations require, at a minimum, either the implementation of the IPMVP requirement of establishing savings estimates with field-measured components or a consumption analysis that includes pre- and post-program utility bills from participants. Few of the 2002-2003 program impact evaluations met this rigor requirement.

Conducting net-to-gross adjustments is also important for projecting accurate saving estimates. Across the portfolio of resource acquisition programs, only twenty-one program evaluations (42 percent) included an analysis of free-ridership instead of accepting the default net-to-gross ratio. Less than half (23 of 50) of the program evaluations took free-ridership into consideration when reporting savings, covering only 28 percent of kWh, 29 percent of demand and 45 percent of natural gas savings goals.

Very few of the evaluations examined any of the different types of net-to-gross adjustments that typically act to increase savings estimates. For example, only one of the reports included the impacts associated with the program changing the way the market operates. Other positive net-to-gross components occasionally measured were participant spillover (2 reports), non-participant spillover (2 reports) and education and information influence (1 report). Where these positive adjustments are not included in the study and free-ridership estimates are, savings estimates may be substantially conservative if the programs had any measurable effects on the operations of the markets in which they are placed - a likely result for larger programs and programs involving education, advertising or training.

Many of the evaluation methodologies, especially for the smaller programs, were hampered by evaluation budgets that were too low to support rigorous or reliable estimates of net energy impacts.

Most studies were found lacking in reliability metrics reported. It is just as important to provide the data needed to understand the reliability of the estimates as it is to provide the estimates. As such, all impact evaluations should report a minimal level of these metrics, including estimate confidence levels, standard deviations and potential bias assessments. Few evaluations provided this information.

If the post-2005 draft evaluation protocols (currently in the public review process) are approved, future evaluation summary studies can report the confidence levels around evaluation-reported program saving estimates and provide a comparative discussion of the potential sources of bias and an assessment of the reliability of the estimates. These metrics were not required and so are not reported in the evaluations of the 2002-2003 Portfolio programs. This will most likely hold true for the 2004-2005 program studies as requirements and budget levels remain similar to those for the 2002-2003 evaluations, although the CPUC took a greater roll in the evaluation planning efforts. The CPUC may want to provide evaluation contractors for the 2004-2005 reports with a set of required reporting metrics to help reduce these problems.

The original RFP and Work Plan for the present project included a significant effort for a summary analysis, comparative analysis and a possible meta-analysis. Yet, the evaluations were too often not comparable and not rigorous enough to support these types of analyses.

The present assessment was also hampered by limited or late data availability. It was significantly delayed by unavailable program workbooks, late evaluation reports and program extensions that acted to delay evaluation efforts. Final program workbooks were repeatedly modified over the course of this assessment and several were missing from the records, requiring additional collection and re-verification work. As of the date of this report, there are five 2002-2003 Portfolio evaluation studies yet to be delivered. It will be important for the CPUC to require evaluation contractors to deliver future studies in a timely fashion.

Lessons Learned and Recommendations

A number of valuable lessons were learned during the course of the present project. The majority can be classified as either related to evaluation implementation and reporting protocols or more administrative, including areas like contracting, budgeting, and planning.

A summary of these and related recommendations include the following:

1. **Develop prescriptive evaluation protocols to guide the evaluation budgeting and implementation processes**. The EEPM was not designed to guide evaluation efforts and tends to be ambiguous in this area. Evaluation contractors and implementers, particularly the TPI contractors, clearly would have benefited from help in understanding the evaluation requirements of the EEPM. This was especially true for complying with IPMVP field procedures, including that all IPMVP options require some form of technology-based on-site field measurements (field surveys and interviews do not comply with IPMVP requirements). Additionally, the budgeting process would have benefited from a implementers having a clearer understanding of the complexity of the evaluation process and what resources are required to provide reliable evaluation results. While providing assistance in these areas is one option, a better approach is to abandon the use of the EEPM as an evaluation guidance document and adopt protocols specifically designed to do so.
2. **Establish clear definitions of the evaluation metrics required from the evaluation process, including those for energy savings (kWh and therms), demand impacts (kW), lifecycle or lifetime savings and evaluation-confirmed net-adjusted savings**.
3. **Establish clear required evaluation protocols, especially for reporting and estimating net ex-post energy impacts and for conducting process evaluations**.
4. **Establish impact estimation protocols and sampling requirements.**

   The metrics and protocols called for in 2, 3 and 4 above are critical if energy efficiency programs are to be relied upon to help plan for California's energy needs. Currently, evaluations vary widely in their levels of rigor, comparability and completeness. Policy makers are significantly hampered when evaluations exclude assessments of ex-post kW effects, limit the reporting to only the first-year savings or exclude estimates of the rate of savings degradation (kW, kWh, and therms). Including program-specific free-rider estimates and participant spillover would more accurately report the program's net ex-post energy impacts. If all impact evaluations are required to report these metrics, California policy makers can use these estimates to update forecasts for the amount of energy that can be provided via energy efficiency, resource acquisition and procurement programs. These metrics are especially important now that California has committed to

relying on these types of programs to provide a portion of the energy growth needs of the state's consumers.

Minimum impact evaluation metrics should include:

- First-year net kWh, kW, and therm impacts;
- Net energy impacts for each consecutive year over which savings are expected (kWh, kW, and therms); and
- Actual program expenditures (dollars actually spent to achieve program savings rather than dollars budgeted) less program evaluation costs.

There is also a need to clearly define the term "*net*" so that it meets energy efficiency policy needs and is identical across all evaluations (including market effects studies that must estimate energy impacts from program-induced changes in the operations of energy markets). This may mean that the term "*net*" should be defined as savings adjusted for free-riders, participant spillover and persistence of effects.

5. **Require that all evaluations present the level of precision and the error bounds around estimates**. Have the evaluations contain a discussion of the reliability of the findings as well as the biases embedded within the approach or the analysis efforts and the approaches employed for reducing the influence of the biases.
6. **Provide clear guidance on how to conduct satisfaction assessments** so they lead directly to program improvement recommendations and can be compared across programs.
7. **Consider the use of market effects studies** to help quantify the market changes that are not captured within the program evaluation studies and the energy impacts associated with these indirect program effects. The evaluation field is moving toward assessments that incorporate moving market baseline estimates of technology adoption rates and causes, compared with the results from technology-specific or technology group-specific studies of total market changes, in order to identify how programs impact both the energy consumption of the participant and of the customers within the influenced market. California should continue to explore how these market-based evaluation approaches can be incorporated into the program evaluation process to provide more reliable short-term and long-term assessments of energy impacts.
8. **Clarify the relationship between program implementers and evaluators**. Do not establish evaluation relationships in which those responsible for program implementation and management are also responsible for evaluation budgeting and approach development efforts. There seemed to be misunderstanding for some implementers and evaluators about the term "arm's-length evaluation effort."
9. **Provide definitions of skills required of evaluation staffs**. This would be especially helpful for implementers seeking and assessing evaluators.
10. **Provide adequate time for the evaluation contracting and planning process**. Too often evaluation efforts were hindered by inadequate planning in these areas.
11. **Require evaluators to submit reports on time**. If programs are extended, consider whether evaluation reports are still needed within a reporting window so that portfolio

assessments can move forward. Portfolio evaluation efforts were hindered by late or lacking reports.

The above-listed lessons learned and recommendations led the 2004-2005 Master Evaluation Contract Team to make recommendations for the 2004-2005 evaluation review process. As a result, the CPUC has used the 2002-2003 lessons to modify the 2004-2005 evaluation review process. Though it was not possible to incorporate all of the desired modifications, many have been and significant changes have already been undertaken within the Evaluation, Measurement and Evaluation (EM&V) Plans for the 2004-2005 program evaluations. However, the EEPM and administrator budget planning process still control the majority of the evaluation efforts and approaches.