EXHIBIT NO. ___(RWS-7)
DOCKET NOS. UE-111048/UG-111049
2011 PSE GENERAL RATE CASE
WITNESS: ROBERT W. STOLARSKI

BEFORE THE
WASHINGTON UTILITIES AND TRANSPORTATION COMMISSION

WASHINGTON UTILITIES AND TRANSPORTATION COMMISSION, **Complainant,** **v.** **PUGET SOUND ENERGY, INC.,** **Respondent.**	**Docket No. UE-111048** **Docket No. UG-111049**

SIXTH EXHIBIT (NONCONFIDENTIAL) TO THE PREFILED REBUTTAL TESTIMONY OF ROBERT W. STOLARSKI ON BEHALF OF PUGET SOUND ENERGY, INC.

JANUARY 17, 2012

research/into/action inc

Final Report

Effective Evaluation Organization Research Report

Funded By:

PSE PUGET SOUND ENERGY
The Energy To Do Great Things

Prepared By:

research/into/action inc

Dulane Moran, MPA
Joe Van Clock
Kara Crohn, Ph.D.
Jane S. Peters, Ph.D.
Research Into Action, Inc.

February 28, 2011

ACKNOWLEDGEMENTS

The evaluation team gratefully acknowledges Laura Feinstein of Puget Sound Energy for her project management support. We would also like to thank the members of Energy Efficiency Services that offered their insight, especially Syd France, Bill Hopkins, Dan Anderson, David Landers and Jeff Trip. Together with Laura, these members of the EES staff participated in a collaborative review process and developed a set of decision considerations for the EES management team.

We would also like to acknowledge David Nightingale of the Washington Utilities and Transportation Commission for his role as project co-host, and thank the members of Puget Sound Energy's Conservation Resources Advisory Group, a WUTC stakeholder committee, for responding to our requests for interviews and offering their perspectives and insights.

Finally, this research would not have been possible without the cooperation of contacts at comparison organizations, all of whom took time to answer our questions and provided detailed information about how they structured and managed evaluation activities at their organizations. We would specifically like to thank Bruce Folsom at Avista Utilities; Fred Gordon at Energy Trust of Oregon; Kevin McKinley of San Diego Gas & Electric; Jennifer Meissner from NYSERDA; Lisa Shea at NSTAR; and Wayne Williams of the Pennsylvania Public Utility Commission.

TABLE OF CONTENTS

E EXECUTIVE SUMMARY

INTRODUCTION AND PROJECT BACKGROUND

In July 2010, Puget Sound Energy (PSE) contracted with Research into Action to investigate and provide recommendations to inform efforts to strengthen the existing evaluation function housed in Energy Efficiency Services (EES). This project is co-hosted by a representative at the Washington Utilities and Transportation Commission (WUTC), although the project was not requested or required by the WUTC.

This is an important project, and the research team would like to acknowledge the PSE staff and management team for undertaking an internally-focused research project designed to guide organizational decision-making. In considering organizational structure and planning decisions, it is important to note that there is rarely one "right" answer. Ultimate decisions about the organizational structure and internal capacity of the EES evaluation unit are the purview of the EES management team and, as such, reflect a host of considerations outside the scope of this research.

APPROACH

As part of this project, the Research into Action team was asked to interview internal stakeholders (members of EES staff) and external stakeholders (members of Puget Sound Energy's Conservation Resources Advisory Group, a WUTC stakeholder committee) and to review the evaluation function at six other organizations engaged in energy efficiency program administration. Research into Action completed these interviews in August and September of 2010. The results of this primary research are contained in this document.

After preparing this report, the research team developed a set of considerations for the EES management team. The Considerations document was reviewed and edited through a collaborative process that engaged members of the research team and a sub-group of EES staff. After reviewing the final Considerations document, the EES management team established next steps and requested an Action Plan to guide and inform a variety of tasks expected to strengthen and provide cohesion for the evaluation team at PSE. That Action Plan is a separate document.

FINDINGS

Internal Stakeholders

EES evaluation stakeholders report that the quality of the evaluation work conducted at PSE is high and expressed no concerns about the credibility of evaluation products currently produced.

There are broad areas of agreement among EES about the quality of the evaluation analysts and the credibility of PSE's evaluation products. Internal stakeholders do not have a common view about the best place in the organizational structure for evaluation. There may be opportunities to improve communication and interaction between the evaluation and program implementation groups.

External Stakeholders

External stakeholders agree on the overall purpose of evaluation, but possess a variety of perspectives on issues of methods, cost, frequency, the role of third-party contractors, use of evaluations, and organizational reporting structures. While their opinions do not diverge too far from each other, they do not demonstrate relative consensus. These contacts offered general criteria for evaluators to follow, but stopped short of advocating specific targets (such as a percentage based allocation to evaluation or minimum frequency).

External stakeholders had few concerns with the overall credibility of PSE's evaluations, but described limited exposure and understanding of the details behind PSE's evaluation work. These contacts reported different levels of understanding about the kind of evaluation work recently completed, the extent to which evaluation findings and recommendations have been used, and how evaluation projects are prioritized.

External Entities

The research team interviewed contacts at six comparison organizations. These organizations, listed in Table ES-1, included three investor-owned utilities serving electric and natural gas customers, two non-utility program administrators, and one Public Utility Commission (PUC). To allow for a more complete comparison, we gathered the same information about PSE.

Table ES-1: Comparison Organizations

Organization	Type
Avista	Investor-Owned Utility
Energy Trust of Oregon	Non-Profit Program Administrator
NSTAR	Investor-Owned Utility
New York State Energy Research and Development Authority (NYSERDA)	Public Benefit Corporation
Pennsylvania PUC	Public Utility Commission
San Diego Gas & Electric (SDG&E)	Investor-Owned Utility
Puget Sound Energy	Investor-Owned Utility

Contacts at all of the comparison organizations valued evaluation results, although their experience managing and the overall level of control maintained over evaluation projects varied

substantially. Comparison organizations faced similar methodological and research tradeoffs—budgetary limitations that limit the ability to achieve a minimum 90/10 confidence and precision; using evaluation resources to resolve uncertainty; and generally prioritizing impact evaluations over process evaluation and market studies.

Maintaining the credibility of their evaluation products is important to all of the comparison organizations. The most common strategy for ensuring a credible product is hiring expert external evaluators to conduct important evaluations. Another approach is to create a process through which stakeholders or third-party advisory groups review evaluation plans or products. This approach works best when the advisors have some technical expertise or evaluation experience sufficient to judge the quality of the work.

Conflict over evaluation results is common, and comparison organizations have established a variety of processes for reducing the impact of these disagreements while retaining the independence of the evaluation process. The most common strategies involve presenting interim and draft findings to program staff and creating formal review or evaluation response documents.

NEXT STEPS

After summarizing the findings from this document, the research team worked closely with EES staff to develop a working Considerations document designed to inform potential EES management team decisions. The six consideration categories were:

1) How evaluation spending is prioritized
2) The level of evaluation expenditures
3) The precision, scope and focus of evaluation work
4) The organizational fit for evaluation team members
5) The integration of evaluation activities and products into program management
6) Integrating evaluation results into rates and regulatory incentives

Review and discussion of this document by internal and external stakeholders informed the discussion and influenced several decisions expected to sustain a highly skilled, professional evaluation unit. These decisions resulted in a separate action plan, developed in December 2010 and January 2011.

1 WHAT IS EVALUATION?

INTRODUCTION

In this chapter we provide an overview of our perspective as to what evaluation is broadly and how evaluation is applied to energy efficiency programs and portfolios. This perspective sets the stage for the guiding principles for how Puget Sound Energy (PSE) evaluation group should move forward. In our interviews with internal and external stakeholders, we uncovered a general lack of knowledge and understanding about program evaluation. We hope that this brief overview provides a context for our recommendations and expands the readers' understanding of program evaluation.

EVALUATION OVERVIEW

Evaluation is an effort to do what individuals and groups do on an everyday basis – gather information to determine which way we should do something – but do it in a more systematic, thorough, useful, and less biased manner. There are a variety of ways in which we make decisions, ranging from relying on individual intuition to employing group discussions to using a formal research processes and findings.

On the research end of the spectrum, evaluations apply social research methods to address questions that are relevant to a particular context. These evaluation contexts are commonly lumped into two categories: formative (e.g. process) evaluation, and summative (e.g. impact) evaluation. Whereas summative evaluations are judgment oriented, "aimed at determining the overall merit, worth, significance, or value of something" (Patton, p. 113),[1] formative evaluations are improvement oriented, seeking to identify strengths, weaknesses, best practices, opportunities, efficiencies, and alternatives for program development, implementation, and replication. Formative evaluation approaches are often applied in situations where continuous improvement and organizational learning are prioritized.

These terms are useful in that they distinguish the purpose of the evaluation and the roles an evaluator may play. A formative or process evaluation might examine the way in which a program is administered and managed, whereas a summative evaluation would likely focus on the impact of the program, including intended and unintended consequences. Traditionally in the evaluation field, these two kinds of evaluation are conducted separately, which helps to minimize bias on the part of the evaluator. In theory, an evaluator engaged in helping a program improve

[1] Patton, M. (1991). "Utilization-Focused Evaluation: 4th edition". Also see definitions in the chapter "Michael S. Scriven: The Science of Valuing" in "Foundations of Evaluation: Theories of Practice" by William Shadish, Jr., Thomas D. Cook, and Laura Leviton (1991).

its planning and administration processes may be too invested in helping the program succeed to recognize some of the unintended outcomes. Energy efficiency program evaluation has tended to encourage joint impact and process evaluations for cost savings and to encourage integration of findings and data collection activities; yet some organizations, such as NYSERDA and the California Public Utility Commission (CPUC) explicitly separate these two activities.

The downside to defining evaluation as formative and summative is that program personnel and evaluators may tend to separate the way in which a program is run from the program's results when, in fact, they are integrally related. For example, an impact evaluation plan may emphasize demonstrating the extent to which a program meets its preset goals. Unless the evaluation also takes aspects of program implementation into account, it risks misinterpreting the results: programs may not be implemented as planned; program plans may yield undesirable consequences; mediating and moderating effects of program activities can have profound effects on program outcomes; and unanticipated events or outcomes can not only influence the results, but also call into question what a desirable program goal should be. The actual program mechanisms need to be understood (as opposed to only considering the planned program processes). Without this understanding, a final judgment of a program's merit may not be possible and/or program outcomes may be incorrectly attributed or misinterpreted.[2]

Over the past 20 years, professional and academic evaluators have increasingly branched out from the process/impact dichotomy to integrate evaluation into organizational management. Organizations that integrate evaluation into their decision-making processes move away from viewing evaluations as one-time events and instead emphasize integrating the evaluation processes into the organization's ongoing decision-making frameworks and building evaluative thinking into the culture of the organization.[3]

Many organizations have embraced the concept of continuous process improvement, and some are beginning to see how evaluation can fit into continuous improvement cycles. The Conservation Measures Partnership[4] is an example of a sector-wide effort to develop and use an open-standards, adaptive management framework for integrating evaluation into decisions about where and how to allocate program dollars. One member organization successfully built a fully-integrated evaluation system using an adaptive management framework. Other members are integrating modified versions of the standards into their management strategies.

---

2 See" Program Theory-Driven Evaluation Science" by Stewart Donaldson (2007)

3 See "Evaluative Inquiry for Learning in Organizations" by Hallie Preskill and Rosalie Torres (1999).

4 The Conservation Measures Partnership website: *http://www.conservationmeasures.org/initiatives/standards-for-project-management*. Link to the Open Standards for the Practice of Conservation *http://www.conservationmeasures.org/wp-content/uploads/2010/04/CMP_Open_Standards_Version_2.0.pdf*

Certain evaluation techniques facilitate this kind of evaluation approach, including the use of program theory models and logic models.[5] In modeling programs, evaluators work with program staff to carefully think through the way in which a program intends to work and the specific steps or processes expected to produce the desired results. A model, such as a program theory, then serves as a basis for comparison and documentation during program tracking:

- Is the program implemented as designed?
- Do the links in the program theory work as intended or in different, unanticipated ways?
- Where are the places in the program process that hinder and help the program's success?

The program theory and logic model also serves as a way for evaluation to test hypotheses and counterfactuals, unearth positive and negative unintended outcomes, demonstrate program progress, demonstrate causal and correlative links between activities and outcomes, and explain how the results occurred. Furthermore, the program theory can be a "living" document, evolving and thus being used to document changes in priorities, goals, or activities.

Regardless of the evaluation approach, the purpose of conducting evaluations is to provide accurate, timely information in a useful and usable manner. If an evaluation fails to do this, then, at the very least, time and money is wasted; and worse yet, decisions could be made based in part on a lack of information or inaccurate information. While this is not the sole responsibility of evaluators, evaluators should work closely with stakeholders, especially the primary intended users of the evaluation, to ensure that some basic standards of quality are met.[6] All involved in the evaluation must make an effort to communicate effectively throughout the evaluation and program planning process; from program inception to conclusion and from evaluation plan to implementation of recommendations.

The Program Evaluation Standards[7] of utility, feasibility, propriety, accuracy, and evaluation accountability provide a good starting point. Coordination with stakeholders requires gaining an understanding of the various contextual factors that are important to a program's success, including political, market, regulatory, interpersonal, economic, and technical factors. There is extensive literature that discusses the evaluator's role, ranging in topic from the way in which evaluators interact with stakeholders, to the values evaluators bring to their work, to the necessity of not only acting as methodologist, but as facilitator, critic, and advisor to program development.[8]

---

5 See "Program Theory-Driven Evaluation Science" by Stewart Donaldson.

6 See Patton, M. (1991). "Utilization-Focused Evaluation: 4th edition".

7 See "The Program Evaluation Standards" published by the Joint Committee on Standards for Educational Evaluation: http://www.jcsee.org/program-evaluation-standards

8 For good examples of actual evaluations discussed in terms of the evaluators' roles and evaluation purposes and uses, see "Evaluation In Action" by Fitzpatrick, Christie, and Mark (2009).

The more an evaluator's role is viewed as part of the overall operation of an organization, the more smoothly and effectively evaluations will run. The evaluator role likely needs to change from project to project in order to adapt the evaluation to the needs of its users. In addition to establishing the evaluator's role on a given evaluation, staff and managerial roles and responsibilities in the evaluation must also be defined. As such, it is a good idea to establish the role at the beginning of the evaluation process. All involved should attempt to adhere to their contracted roles unless those roles need to be formally renegotiated during the evaluation process.

Building a successful evaluation system is facilitated in part by establishing and maintaining trust and credibility among evaluation staff, program staff, the program manager, upper management, and external stakeholders. However, these stakeholders must also be invited to co-construct the evaluation system in a way that meets organizational needs while attaining evaluation standards of practice. This may require some learning on the part of all stakeholders to understand the value of evaluation to programmatic and organizational decisions and develop the capacity to think in evaluative or inquiry-based ways[9].

An important aspect of achieving a robust evaluation system is understanding the multiple ways in which an evaluation will, or could, be used by stakeholders. For example, an evaluation of a program may be used not only by program managers to institute programmatic changes, it could also be used in conjunction with other program evaluations and planning materials to make decisions about a portfolio of programs or efforts to organize regional or statewide programs.

In the case of energy efficiency evaluations, the evaluators must understand how programs fit with a provider's overall portfolio of programs, as well as the regional markets so that the evaluation design feeds into the integrated resource planning (IRP) and a larger understanding of energy efficiency efforts. There is a deep literature on factors influencing the usefulness of evaluations, including various ways in which the evaluation process itself can be used to assist program and organizational development.[10]

ENERGY EFFICIENCY PROGRAM EVALUATION

Energy efficiency program evaluation is a specific environment in which evaluations are conducted. Two documents that provide a useful framework for thinking about energy efficiency program evaluation are the *National Action Plan for Energy Efficiency* (NAPEE or *National Action Plan*) and the *California Evaluation Framework (Framework).*[11]

---

9 See "Evaluative Inquiry for Learning in Organizations" by Hallie Preskill and Rosalie Torres (1999).

10 See Alkin, M. & Taut, S. (2003). Unbundling Evaluation Use. *Studies in Educational Evaluation,* 29, 1-12 for an explanation of various forms of process use.

11 The Leadership Group (2006). National Action Plan for Energy Efficiency. United States Environmental Protection Agency & United States Department of Energy, Washington DC. TecMarket Works Team.

continued...

The National Action Plan defines program evaluation in this way:

> *Program evaluation informs ongoing decision-making, improves program delivery, verifies energy savings claims, and justifies future investment in energy efficiency as a reliable energy resource. Engaging in evaluation during the early stages of program development can save time and money by identifying program inefficiencies, and suggesting how program funding can be optimized. It also helps ensure that critical data are not lost." (*pp. 6-45)

The *Framework* was prepared for the CPUC. The document is designed to provide "*a consistent, systemized, cyclic approach for planning and conducting evaluations of California's energy efficiency and resource acquisition programs*."(p.1)

Consistent with *The National Action Plan*, the *Framework* defines the purpose of evaluation as "*to help ensure that good decisions are made regarding the investment of energy program resources by providing rigorous, independent evaluation studies and study results."*(p.23)

While the *Framework* targets evaluators of California's energy efficiency programs, it is basically a primer on the various evaluation activities that can be conducted for energy efficiency programs and provides discussions of best practices for methods and approaches for impact evaluation, measurement and verification (M&V), process evaluation, information and education program evaluation, market transformation program evaluation, non-energy effects evaluation, as well as uncertainty, sampling, and cost effectiveness.

The two broad categories of evaluation for energy efficiency programs are effects (also known as impacts) and process evaluation. Effects evaluations include energy impact, M&V, market, and evaluations focused on measuring non-energy effects. Process evaluations document and understand program operations and identify opportunities for program improvement, as well as study market operations and processes and other conditions that affect program operations. Each of these types of evaluations are defined in the Glossary at the end of this chapter.

The *National Action Plan* documents a variety of best practices for energy efficiency program evaluation (p.6-50):

- Incorporating an overall evaluation plan and budget into the program plan.
- Adopting a more in-depth evaluation plan each program year [e.g., a continuous improvement approach].
- Prioritizing evaluation resources where the risks are highest. This includes focusing impact evaluation activities on the most uncertain outcomes and highest potential

---

(2004). The California Evaluation Framework, project number K2033910. California Public Utility Commission, San Francisco, CA.

savings. New and pilot programs have the most uncertain outcomes, as do newer technologies.

- Allowing evaluation criteria to vary across some program types to allow for education, outreach, and innovation.
- Conducting ongoing verification as part of the program process.
- Establishing a program tracking system that includes necessary information for evaluation.
- Matching evaluation techniques to the situation in regards to the costs to evaluate, the level of precision required, and feasibility.
- Maintaining separate staff for evaluation and for program implementation. Having outside review of evaluations (e.g., state utility commission), especially if conducted by internal utility staff.
- Evaluating regularly to refine programs as needed (changing market conditions often require program changes).

Implementing these best practices facilitates energy program administrator's growth in evaluation capacity by not requiring that program administrators have full evaluation capability at the outset of their efforts with energy efficiency program and energy efficiency program evaluation.

2 INTERNAL STAKEHOLDERS

INTRODUCTION

This chapter describes key themes that emerged from in-depth, confidential interviews with 12 internal stakeholders, primarily PSE evaluation and program staff, conducted between July 29, 2010 and August 23, 2010. This document provides topics for immediate discussion and guidance for the other products of this evaluation: specifically a set of considerations developed collaboratively with PSE staff and, ultimately, an action plan for realizing these principles.

SUMMARY

There are broad areas of agreement among Energy Efficiency Services about the quality of the evaluation analysts and the credibility of PSE's evaluation products. Internal stakeholders do not agree about the best organizational structure for evaluation and expressed a desire that improved communication and a higher level of professional interaction in general guide interactions between the evaluation group and the program implementation group. Inter-group conflicts may be a vestige of past conflicts, but are aggravated by a lack of experience with energy efficiency, disagreements over evaluation findings or methods, and conflicting perceptions about the degree and manner in which evaluation products should or can be used.

FINDINGS

Purpose of Evaluation

According to internal stakeholders, impact evaluation is the first priority, but process evaluation and market assessments are also valued, particularly by program managers seeking ways to improve their programs. All contacts agree that at PSE, the primary purpose of evaluation is to provide an estimate of the amount of kWh savings generated by energy efficiency programs and to compare that estimate to the pre program savings assumptions to attain a realization rate.[12] If an evaluation reveals a realization rate less than one, then stakeholders expressed an expectation that the evaluation offer explanations for the difference and suggestions for either program adjustments and/or adjustments to savings estimate algorithms.

[12] Two internal stakeholders used the term 'verifying savings' in their statement on the purpose of evaluation. This could just be terminology, but it suggests a lack of specificity in understanding that verification is primarily observation that measures have been installed as claimed, not a means for assessing whether savings are being achieved as claimed.

There were fewer overall comments about process evaluations at PSE. Contacts affiliated with evaluation noted they occasionally occurred. Program management contacts provided more extensive comments about the potential value of process evaluations, particularly as they would provide information about the effectiveness of specific program components or strategies. For them, process evaluations can provide information useful for improving program delivery.

According to internal contacts, the market research group typically conducts market studies, some of which might be market evaluations. The level of information in the market studies tends to be at a high level—as opposed to the program-specific or measure-specific information obtained through evaluations. One contact noted that there may be a lack of consensus about the purpose of market studies.

Planning and Program Adjustments

Internal stakeholders were specifically asked to discuss the role of evaluation staff in program design, planning, implementation and evolution, including how evaluation products are used. Contacts saw evaluation as a support role, providing information or interpreting information important to program staff at strategic points in a program's life cycle. Program managers then use that information, combined with their own understanding of program operations, to determine a course of action and inform program design decisions.

Internal stakeholders agreed that there was value in having evaluation staff involved in program planning discussions. Contacts offered several ways in which this was valuable: it ensures the evaluation plan will support the program as planned; it allows evaluators to provide insight based on their knowledge of other programs' performance; and it increases the likelihood that the necessary data would be collected during implementation. There were indications that communication and coordination during the planning process was evolving and recent efforts had been inconsistent. Evaluation staff might be asked to provide estimates of program savings potential based on a specific design or group of measures, or they may be unaware of a pilot program or inclusion of a new measure until after the program planning had occurred.

Responses indicated a lack of consensus about how and when evaluation staff should be involved in program design. On one hand, contacts saw value in having evaluation staff engaged enough to understand the programs and provide insight. On the other hand, some contacts were quite clear in their desire to limit the involvement of evaluation staff in design, arguing that engaging evaluation staff in program design could create a conflict of interest when that design was up for evaluation or that a limited role in design and implementation was appropriate, given the importance of professional independence.

All internal stakeholder contacts reported valuing on-going communication during evaluations. However, it was not clear that contacts had the same expectations or were considering the same experiences when discussing how this communication occurred. Program staff contacts provided examples where evaluation projects occurred without sufficient interaction, and that this limited the usefulness of the evaluation findings.

One of the more complicated disagreements emerged when contacts were asked to discuss how evaluation results are used. Evaluation staff believe their work is valuable and that a measure of that value is the extent to which it is used by program staff. Other contacts, including those affiliated with programs; believe evaluation products are one of many valuable inputs in their program design and management decisions. For these contacts, other considerations might overshadow evaluation recommendations, or limit the speed with which recommendations are incorporated. One contact provided a suggestion for navigating these conflicts; this contact wanted to see evaluation staff work collaboratively with program staff to identify strategies for incorporating evaluation findings.

Evaluation Approach, Cost, and Frequency

Approach

In terms of evaluation approaches, few outside of the evaluation staff expressed opinions on specific approaches or methods chosen for the evaluations themselves. That said, there is a clear desire on the part of some evaluation staff and many program managers to pursue more market assessment and process evaluations that would better inform program design and implementation process decisions; evaluations that explore not only what kinds of incentives and rebates to offer at what price point, but also how to manage day-to-day operations in the most effective way. Contacts discussed several methods for determining cost savings and decision-making processes, other approaches to evaluation were not discussed. Specifically, no contacts indicated an awareness of evaluation approaches that use evaluation activities to complement the utility's efforts to create continuous improvement mechanisms, such as utilization-focused, participatory, collaborative, and program theory-driven evaluation approaches.

Decisions about methods are currently the purview of evaluation staff members, who report that selecting an evaluation approach and appropriate metrics is their responsibility. Contacts affiliated with programs did not dispute this, but noted that they would be better able to support successful evaluation if they were involved in evaluation planning discussions. This desire emerged primarily from concerns about data requirements: contacts described scenarios in which data that were needed had not been collected as well as data that had been assumed to be unattainable were, in fact, available. For one contact, the overriding consideration for evaluation methods was that the results be statistically valid.

Costs

While expectations for certainty and statistical validity did emerge in discussions of evaluation methods, they did not emerge as a driving consideration for evaluation costs. Statistical validity is tied to sample size and expected variation and can thus increase costs as the volume of data required grows. In their discussions with us, none of the internal stakeholders advocated for an evaluation budget tied to a percent of the program portfolio costs. Instead, they sought flexibility, stating that evaluation costs should be determined based on:

- Complexity of the measure or project
- Timing of program changes
- Status as pilot program
- Perceived controversy associated with measures, program approach or market conditions
- Uncertainty in kWh savings estimates

Frequency

Mirroring their opinions on evaluation costs, internal stakeholders advocated for case-by-case consideration when determining an appropriate expectation for the frequency of evaluations. Many of the factors that might affect evaluation cost would also affect how frequently a program or measure should be assessed: how rapidly a program is changing; how stable the realization rate is; indications of market changes; or the presence of new measures. Suggestions for evaluation frequency ranged from an annual review for some programs to only after substantial program changes, typically every four or five years.

Credibility of Internal Evaluations

Internal stakeholders generally reported being satisfied with the credibility of internal evaluation work but expressed concerns about the perceptions of others, particularly members of the CRAG. These concerns emerged not from any direct comment about PSE's evaluation work. Rather, contacts were aware that other utilities had experienced issues with evaluation quality and that at PSE the evaluation team reported to the same director as the program implementation team, potentially opening the evaluation staff to accusations of conflicts of interest.

Credibility is driven by two fundamental facets: quality of work and perception of neutrality. Contacts outside of the evaluation group were generally satisfied with the level of work quality and believe the evaluation staff members have the ability to conduct accurate and appropriate analyses. Several of these contacts believed evaluation staff would be more effective if they had a fuller understanding of the programmatic or market issues that can affect the evaluation design or usability of findings. Staff comments on evaluators' neutrality were nuanced. Some comments reflected experience with specific staff members having preconceived notions or beliefs that a project was not conducted appropriately. For the stakeholders reporting these concerns, achieving neutrality remained an area in need of improvement.

Evaluation staff members believe their work is viewed as credible most of the time. When program staff members do question the evaluations, the evaluators view the criticism as rooted in program managers' disappointment with the results more than true concerns over decisions about the evaluation approach, methods, or analyses. Members of the evaluation group designed a process (the Evaluation Response Report) to help navigate discussions about evaluation results

and recommendations to reduce emotional responses that can interfere with decisions about making program changes. In general, staff reported that this process seems to be effective.

Group Dynamics

Within-Group Conflicts

Contacts were aware of conflicts that occurred within the evaluation team. We learned that the evaluation group is conflicted about how evaluation staff should interact with program staff and that there may be disagreement about the appropriateness of a reporting structure that locates the evaluation function under new program design within program implementation. Finally, some tension may be resulting from a perceived lack of management support; particularly when evaluation staff find themselves having to navigate contentious inter-group relationships and defend their work to the program group.

Inter-Group Understanding and Conflicts

The relationships between the evaluation team and managers of non-program divisions are relatively effective. While they exchange information and generally hold each other in high regard, there is less frequent and less direct contact among these groups than there is among evaluation staff and program staff. The managers of other divisions have an incomplete understanding of the nature and scope of the evaluation team's work.

Inter-group conflicts are felt most deeply and occur most often between the evaluation group and the program group. Some of this conflict is the result of adversarial relationships between these groups in the past. Contacts described an environment in the past in which there seemed to be the view that it was the program managers' job to set up the programs and evaluators' job to shoot them down. Although that tension has subsided as evaluators and program managers work more closely together throughout the program design and implementation process, any lack of trust in each other's intentions becomes problematic when evaluation results are contentious.

Evaluation team members and program staff members also acknowledged that personality differences can create problems. Program staff members understood the evaluation teams' concerns about the importance of accurate data collection and ensuring that programmatic changes occur in a timely manner but these contacts also wanted the evaluation group to gain understanding of and empathy for program operational challenges and work more collaboratively with program staff. Many contacts wanted to move beyond the adversarial relationship and build a process in which evaluation is used as a tool for program improvement as well as estimation of savings. Others sought more distance between evaluation and program staff in order to maintain the evaluation team's independence.

To facilitate better relationships, several managers expressed a need for better definitions of roles and responsibilities; a better understanding of who is in charge and where to turn for guidance at various points in the program development and implementation process and the evaluation

process. For example, during program development, the program manager is the decision maker and the evaluator offers guidance. During evaluation design, the evaluator is the decision maker and the program manager offers guidance.

In addition to defining roles, several contacts expressed a hope that evaluation staff and program staff will learn to interact with higher levels of professionalism. Increased professionalism was viewed as a solution to unproductive or hostile communication, an unwillingness to discuss alternatives in evaluation planning, and evaluation products that are not used or useful. In some cases increased professionalism was tied to increased knowledge or understanding about the energy efficiency field in general. In others, it was tied to management expectations for productive communication habits.

Organizational Structure for the Evaluation Function

Our interviews revealed no internal consensus on where evaluation should exist in the organization with some contacts holding no opinion on the matter.

Three schools of thought emerged in discussions of organizational structure:

1. While there were no internal concerns about the credibility of evaluation products, several contacts expressed concerns about how the CRAG and WUTC might perceive the independence of the evaluation group because it reports to the same director who manages program implementation.

2. There were comments that reflected the belief that the evaluation group should report to a different manager than the Director of Customer Energy Management both to obtain more managerial support for the evaluation group and to separate the evaluation group from the function of program development and program implementation.

3. Other contacts advocated for keeping the evaluation group directly under the Director of Customer Energy Management to facilitate better collaboration during the program development, implementation, and evaluation processes.

3 EXTERNAL STAKEHOLDERS

INTRODUCTION

The chapter presents the key themes that emerged from in-depth, confidential interviews with nine external stakeholders conducted between August 4, 2010 and September 3, 2010. The external stakeholders were all members of the Conservation Resource Advisory Group (CRAG), an advisory group that works with PSE on development of energy efficiency plans, targets and budgets. The CRAG consists of ratepayer representatives, regulators, and energy efficiency policy organizations. One of the contacts was new to the CRAG, and was unfamiliar with PSE and with energy program evaluation. At the time of the interview, he had not yet formed opinions on these issues and he comments are not included in the discussion below.

SUMMARY

While external stakeholders hold similar views on the general purpose of evaluation, they approach issues of methods, cost, frequency, third party contractors, use of evaluations, and organizational reporting structures from a variety of perspectives. Their opinions do not diverge too far from each other, nor do they demonstrate relative consensus. They offered general criteria for the evaluation team and evaluators to follow, however few felt it necessary to specify targets, such as evaluation budgets as a certain percent of a program budgets or the frequency of evaluations.

Most stakeholders expressed little concern with the credibility of PSE's evaluations, but they lack a broad understanding of PSE's evaluation work. It is clear that they do not share a common understanding of the kind of evaluation work PSE has done to date, the extent to which and ways in which evaluation findings and recommendations have been used, or the evaluation project prioritization process.

KEY THEMES

Purpose of Evaluation

In conversations with external stakeholders, all of whom were members of the CRAG, it became clear that I-937 was affecting the expectations and relevance of evaluation. By requiring utilities to obtain all cost-effective energy conservation, I-937 has increased the focus on verifying acquisition and on the cost-effectiveness calculations themselves. Although many CRAG members do not read evaluation reports in detail, they reported that evaluation results guided their decisions and emphasized the important role evaluation should play at PSE.

CRAG contacts identified four overarching purposes for evaluation:

- Estimate the energy savings from energy efficiency measures and programs
- Confirm the cost-effectiveness of energy efficiency programs
- Improve program effectiveness
- Identify emerging strategies and adapt programs to changing market conditions

CRAG contacts differed in their descriptions of the relative importance of different evaluation strategies. Three of the nine contacts specifically valued impact evaluations for their ability to provide inputs to assess the cost-effectiveness of programs. However, even these contacts recognized the need to look more broadly at market and program processes to understand the full picture. Two other contacts placed primary importance on using evaluation to support program improvement and saw process evaluations as part of a continuous improvement effort. Two others placed equal importance of estimating post installation savings and supporting program improvement, stating that process and impact evaluations should be given equal weight. Few external stakeholders reported seeing or knowing of market or process evaluations conducted by PSE.

Evaluation Approach, Cost, and Frequency

Coordination with Program Planning, Design, and Implementation

The external stakeholder contacts offered a variety of opinions on ways in which evaluation staff should coordinate with program planning, design, and implementation. The range in responses reflects the external stakeholders' views on the extent to which evaluation primarily plays a role for verification as opposed to playing a more involved role of continuous improvement for the programs. This section describes a few issues that approximately half (four to five) of the contacts agreed upon as well as some issues raised by individuals.

Regarding evaluation designs, four contacts stated that program staff should offer advice to evaluators; perhaps help evaluators develop evaluation questions because of the program staff's industry expertise. However, these contacts felt that because evaluators are experts in measurement, evaluators should say what needs to be measured and make sure it is done in an effective way. One contact with evaluation experience felt that, although impact evaluations ultimately demonstrate whether or not a program meets its guidelines, evaluators must do process evaluation in order to make impact evaluation possible. This, the contact says, is because the two kinds of evaluations are interconnected. That said the contact felt that the balance of impact versus process evaluation will vary depending on the program. A few other contacts also understood the importance of matching the evaluation design to the program's stage of development and information needs. For example, one contact pointed out how a process or market evaluation could identify how a program that would otherwise be cost effective could be inefficient in the wrong hands.

In terms of appropriate methods, few contacts offered specific advice. Some contacts felt it was important to follow industry standards and best practices. One contact recommended, in the case of billing analysis, that evaluators be careful to take into account a variety of factors that can affect analyses. For example, measures may show increases in energy use in the short term but will yield longer-term energy savings, e.g. installing a device that was not present before but may help lower energy costs over the long term. Another contact noted that he understood deemed savings numbers can be inaccurate and, as such, would not expect evaluators to analyze data at too granular of a level. Yet another contact recommended that more evaluation work should ask questions about how to increase penetration rates, identify adequate incentive levels, and explore different messaging to go with measures.

During program design, planning, and implementation, five contacts emphasized the importance of early and frequent communication between the evaluation and program staff in order to ensure proper data collection. One of these contacts emphasized the importance of collecting data that would be used rather than collecting data for the sake of simply having data. Five contacts suggested that when a third party evaluation consultant is involved, both internal and external evaluators should contribute to the evaluation plan and offer considerations for program planning.

Contacts view the independence of the evaluation team as an important issue, but see it from a variety of perspectives. Five of the contacts see the need to balance coordination among the teams with evaluator independence, ensuring the evaluation team is able to remain distant enough from program teams to control biases. How that is accomplished is less clear. For example, three contacts felt the evaluation staff should have little to no role in program implementation where as two other contacts did not distinguish evaluation's role in planning versus implementation.

In terms of reporting structure, their opinions ranged. Two external stakeholders said it is important for the evaluation team to report to a different boss than program staff and one thought the evaluation team should be somewhat separate from program staff. Another external stakeholder had no problem with the evaluation staff reporting to the same person as program staff as long as the organization functioned collaboratively. This person felt the reporting structure was less important than the organization's culture in maintaining independence and good relations. Three other contacts had not formed an opinion about the reporting structure.

Use of Evaluation Findings and Recommendations

All external stakeholders felt program managers and staff should use evaluation findings and recommendations. Many offered cautions and advice on how results should be used.

Four contacts suggested that it would be a waste of time if program mangers and staff did not seriously consider evaluation recommendations, especially regarding cost effectiveness. These contacts agreed that evaluators should be at the table to listen and help clarify program goals, metrics, and outcomes. They also said evaluators should come back to discuss findings and offer

feedback in a timely manner. However, these contacts cautioned that evaluation recommendations may not take into account a multitude of factors. For example, evaluation staff do not necessarily have the subject expertise to design programs and, as such, program staff should not solely rely on the evaluators' design recommendations.

Other contacts went on to state that while evaluators should make recommendations, they should not be in a strong enough position to directly dictate what a program does or hold indirect power over program staff. These contacts felt that evaluators can offer good savings estimates for design but should not be so involved as to compromise their independence. One contact noted that sometimes a program's savings estimates do not match the result, but that program may need time to establish itself before making adjustments. Another contact expressed the importance of differentiating between recommendations that could be addressed without CRAG or Commission approval, for example a mid-season addition of a new technology to a program's offerings, versus more substantial decisions requiring external approval.

Cost and Frequency

Most think the cost and frequency of evaluations should be determined on a case-by-case basis given the complexity of issues, timing of program changes, and changes made since the previous evaluation of a program. Although most individual contacts did not feel there was a need for a minimum evaluation budget, they thought that other members of the CRAG supported having such a budget. A few advocated against setting "false minimums," which might lead to wasting money. Most external stakeholders seemed to assume that a minimum budget was important to the other external stakeholders when, in fact, it was only marginally important to a few. Those few members who did suggest a minimum overall evaluation budget did not have strong opinions on the amount and indicated that they were largely guessing at what the "right" amount should be based on other utilities' budgets or industry standards. The suggested range was from one to four percent of program portfolio budget.

The external stakeholders do hold strong opinions on the criteria by which the evaluation and program staff should determine the frequency and costs of program evaluations. The following lists express the range of criteria offered by the external stakeholders.

The cost and frequency of an evaluation should:

- Depend on the size and scale of the program, emphasizing the most important programs. Not all evaluations will be "Cadillac" in quality or in confidence levels.
- Not waste time, but do a good job because the program is important.
- Be sufficient to confirm that incentive levels are appropriate and targeted appropriately to achieve the best performance and best savings.
- Be done frequently enough to capture shifts in the cost effectiveness of programs.
- Reflect the importance of the program to rate payers.
- Be sufficient to ensure that key drivers of a program are demonstrated; the importance of some drivers are not apparent in the short term, but emerge strongly over time.

- Take into account instances in which evaluations may cost disproportionately more at first and less as it goes.
- Correspond with the level of risk associated with program cost, uncertainty, or other pressing needs.
- Occur more frequently for newer programs (six months – one year) and less frequently for more stable programs (every three years).
- Meet I-937 guidelines on conducting cost effectiveness studies
- Scale appropriately to the needs of the program as one would with an adaptive management approach, paying attention to transforming markets, and questioning assumptions.
- Occasionally look at the whole portfolio to help the CRAG understand what has been evaluated over past four years, what the top priorities are, what recommendations have been implemented, and in what direction the evaluation team is heading.

Use of Third-Party Evaluators

All external stakeholders thought third party evaluators should be employed in some circumstances and with a loosely defined degree of frequency, while their opinions differed on the nature of those circumstances, it is apparent from their comments that they would likely agree with each other on most of these perspectives. Three external stakeholders emphasized the importance of having internal staff with evaluation expertise oversee third party evaluations. These stakeholders felt that internal staff should only conduct evaluations of small programs and hire third parties to evaluate the largest four to six programs, complicated or controversial programs, or evaluations of their operations and administration of programs. These stakeholders also advised hiring a third party when the internal evaluation staff lacks necessary skills to conduct the evaluation.

The other external stakeholders offered a range of reasons and circumstances when third party evaluators should be hired. One contact stated that, if PSE has the skills to do internal evaluations well and independently with sufficient constraints in place to "firewall" evaluation staff from other divisions, then they should keep third party evaluations to a minimum to help keep costs lower. Another contact acknowledged that the appearance of independence that third parties provide is helpful and may be needed in circumstances when incentive mechanisms are tied – to any degree –to evaluation results. This contact stressed that it would be optimum to tie evaluations to program processes, conducted in a collaborative manner to ensure effectiveness.

Two contacts said third parties should be hired periodically to review the portfolio of programs over time. One of these contacts also noted that, beyond lending credibility and an additional, objective perspective to evaluations of large scale, long-term programs, third parties could bring expertise from outside the region to new, experimental programs. Another contact agreed that robust comparisons with neighboring utilities would be helpful, but discussed the need for comparisons across the breadth of programs. This contact emphasized that the need for comparative work was not a reflection on the credibility of PSE's work. Several other external

stakeholders concurred that although third party evaluations can offer additional objective perspectives, they are not concerned about the actual credibility of PSE's evaluations. Only one external stakeholder said more third party evaluations should be conducted in order to improve PSE's credibility.

Credibility of Internal Evaluations

External stakeholders' perceptions of PSE's credibility range from not having formed an opinion to holding positive views of PSE's work. No one expressed strong concerns about PSE's evaluations. Several contacts that are new to working with PSE have not yet formed opinions about PSE's evaluation work, expressing mild concern about the independence of the evaluation work given the reporting structure. Three contacts with more experience working with PSE said they had no reason *not* to trust PSE, but implied they did not give a full vote of confidence.

No one expressed concerns over the accuracy of the evaluation analyses, but one contact indicated dissatisfaction with the lack of process evaluations, saying that impact evaluation is not always the best approach; that the balance between impact and process evaluations was uneven and that the evaluation team should also look at how resources are acquired. One contact noted that many years ago evaluation was not done well, but there are currently no concerns of misconduct. This contact also noted that the new tracking systems and ways of demonstrating savings estimates were a good improvement.

When asked if they had particular expectations for the level of certainty evaluations should achieve, several contacts had no opinion, saying it was not their expertise. Two contacts thought that 95% confidence intervals were standard, and another contact suggested that the 90/10 rule was not always achievable, but wanted to see at least 80% certainty, especially for mass market studies. Two other contacts suggested a more general criterion of trying to achieve what is feasible, but not worrying about achieving specific confidence levels.

Use of Evaluation by External Stakeholders

Some external stakeholders use the evaluations on a regular basis to either inform their understanding of PSE's progress towards conservation targets according to I-937 standards and the utility's ability to run cost-effective programs. Some read the evaluations as part of their broader efforts to stay current on regional energy conservation trends and progress. Some read the evaluation reports to inform their work at their organization. Others do not read the evaluation reports often, but reference them for policy decisions or in preparation for testimonies. It was clear from the interviews, however, that none of the external stakeholders have a good understanding of the scope of work the PSE evaluation staff have done or are pursuing.

Most external stakeholders are aware that their fellow CRAG members use the evaluation reports to support their particular perspective on energy conservation, such as controlling costs for rate payers, pursuing more aggressive conservation goals, and making sure low income households

are treated equitably. One contact was concerned that some members might use the evaluation results selectively to target programs for additional support or for elimination as opposed to advising PSE at the portfolio level. A few suggested that fellow CRAG members use, or ought to use, the evaluations to track the cost effectiveness of programs and PSE's ability to meet I-937 criteria and to advise PSE on ways to improve.

Ways Evaluation Could Better Serve Washington State

Most of the external stakeholders felt that there were opportunities for PSE's evaluations to be used more broadly than solely for PSE's purposes. One contact suggested that PSE's evaluations might raise the bar for evaluations done in the region. Other contacts thought it would be useful if all of the Washington investor-owned utilities (IOUs) engaged in collaborative efforts rather than treating evaluation as proprietary. One contact suggested that sharing results could cut down on evaluation costs and improve region-wide knowledge. Another contact pointed out that collaboration sometimes happens serendipitously, but it should happen in a more deliberate fashion to leverage program designs region-wide or at least better document market effects. This contact also suggested that making the results of impact evaluations that documented savings more widely available could provide more confidence in the results. Another contact warned that while collaboration is desirable, that does *not* suggest that EM&V be standardized across IOUs. One contact noted the Northwest Energy Efficiency Task Force as an example of regional collaboration. A few external stakeholders had no opinion on the subject.

From the contractors' perspective, one contact saw the potential for contractors to better understand the aspects of a program that work or do not work for PSE and then make similar changes to their approach.

Inter-Team and Inter-Group Dynamics

The external stakeholders are generally not aware of the inter-team and inter-group dynamics within PSE. A few contacts are aware of some internal conflict, but are generally too far removed from PSE internal workings to hold an opinion about it. Most have no opinion about individual members of the evaluation staff. A few indicated that they think evaluation team members are skilled analysts.

Two contacts offered general advice for the evaluation team. One contact thinks evaluation staff should have evaluation background, and although it would be good to have evaluation experience in the energy field, the contact felt it was more important for evaluation staff to have a good understanding of evaluation best practices. In terms of analytical skills, the contact suggested evaluation staff should at least have an understanding of cost benefit analysis, even if they are not experts in performing the analyses. The other contact suggested that evaluation staff needs to be attuned to program staff's perspectives and ensure that sensitive relationships are not damaged during evaluation. This contact felt that managers should facilitate collaboration among evaluation and program staff, but ultimately be prepared to make the call if disputes arise.

Organizational Structure for the Evaluation Function

Most external stakeholders do not think there are problems of the evaluation staff maintaining independence from the program staff, but a few contacts think others perceive a problem and therefore recommend that evaluation not report to the manager of programs. One contact is skeptical of the objectivity of evaluation studies regardless of where evaluation staff report or whether or not third party evaluation firms are used. One contact does not have an opinion on where the evaluation staff report, but wants the reporting structure to be transparent.

There are a few contacts with strong views on where evaluation should report. One contact thinks PSE already has staff that is somewhat separate from the program staff, and that this is a good approach. This contact explained some reservations about the evaluation staff having direct ties to program staff, using a hypothetical situation: "if a group used to be heavily involved in implementation and a few [members of that group] got pulled out to do evaluation, is that truly independent from implementation? Not really." Another contact recommends that evaluation have more independence from program staff and, as such, should report to the VP of Energy Services or to the Manager of Budget and Administration, not to the Director of Customer Energy Management or the Manger of New Program Development and Evaluation.

Another contact countered that there does not need to be so much concern over the independence of the evaluation team, explaining that if there is a good feedback loop for evaluation to interact with program design, and that it is part of the culture and organizational structure to work collaboratively, then there is no issue.

4 EXTERNAL ENTITIES COMPARISON ANALYSIS

INTRODUCTION

In an effort to provide PSE with additional context for making decisions about organizing for evaluation, we present information about six other organizations engaged in energy efficiency program administration. Working with PSE and WUTC staff, we selected a diverse set of organizations that provided energy efficiency services to both electric and gas customers, most of whom operate with specific evaluation guidelines or statewide protocols. The examined organizations include three investor-owned utilities serving electric and natural gas customers, two non-utility program administrators, and one Public Utility Commission (PUC). To allow for a more complete comparison, we gathered the same information about PSE. Table 1 lists some key characteristics of each organization.

Table 1: Organizational Characteristics

Organization	Organization type	Ratepayers Served		Service Territory
		Electric	Gas	
Avista	Investor-Owned Utility	355,000	314,000	Washington, Idaho, Oregon (gas only)
Energy Trust of Oregon	Non-Profit Program Administrator	1,364,308[i]	646,193[ii]	Oregon
NSTAR	Investor-Owned Utility	1,100,000	267,000	Massachusetts
New York State Energy Research and Development Authority (NYSERDA)	Public Benefit Corporation	6,555,222[iii]	2,356,303[iii]	New York
Pennsylvania PUC	Public Utility Commission	5,628,325[iv]	2,142,844[v]	Pennsylvania
San Diego Gas & Electric (SDG&E)	Investor-Owned Utility	1,400,000	840,000	Southern California
Puget Sound Energy	Investor-Owned Utility	1,000,000	750,000	Washington

[i] Oregon customers of Portland General Electric and Pacific Power.

[ii] Oregon customers of Cascade Natural Gas, Oregon and Washington customers of Northwest Natural Gas.

[iii] Customers of Central Hudson Gas and Electric Corporation, Consolidated Edison Company of New York, Inc.; New York State Electric and Gas Corporation, National Grid; Orange and Rockland Utilities; and Rochester Gas and Electric Corporation.

[iv] Source: 2008 EIA -861 Database. Figure reflects ratepayers of all investor-owned electric utilities in Pennsylvania. Note that only utilities with more than 100,000 customers are subject to Act 129.

[v] Source: Compilation from websites of natural gas utilities regulated by the Pennsylvania PUC.

Our comparison work involved two primary activities. First, we reviewed documents related to energy efficiency programs and evaluation. Second, we conducted in-depth interviews with individuals involved in the organization's evaluation activities. These interviews lasted between thirty and ninety minutes and took place between August 17, 2010 and August 31, 2010. After conducting interviews, we produced a summary document describing evaluation activities at each organization and provided the document to interview contacts for review. Using qualitative analysis software, we then compared findings across organizations.

In addition to the data we gathered from comparison organizations, this chapter draws on a report released in 2010 by The Lawrence Berkeley National Laboratory (LBNL).[13] The report, entitled *Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs*, focuses on evaluation, M&V and verification approaches used in 14 states and by the Northwest Energy Efficiency Alliance (NEEA). The report draws on interviews with more than 50 energy efficiency policy experts, regulatory staff, program administrators, evaluation project managers, and evaluation practitioners.

SUMMARY

Evaluation results were important to all organizations regardless of the presence or absence of protocols. Contacts reported similar considerations in methodological approaches: budgetary limitations that limited the ability of every study to achieve a minimum 90/10 confidence and precision; a focus of resources on resolving uncertainty; and prioritizing impact evaluation work over process evaluation and market studies. PSE allocated the smallest portion of its energy efficiency budget to evaluation and was unique in separating market research and other strategic planning associated with energy efficiency from energy efficiency program evaluation.

All of the comparison organizations seek to maintain the credibility of their evaluation products, but do so in a variety of ways. The most common approach is to hire expert external evaluators to conduct important evaluations. Another approach is to engage stakeholders or third party advisory groups and request that they review evaluation products. This approach works best when the advisors have a level of technical expertise or evaluation experience sufficient to judge the technical quality of the work.

Disagreements or conflict over evaluation results is common and organizations have established a variety of processes for overcoming these disagreements while retaining the independence of the evaluation process. A major component of resolution is the timing as to when

[13] Messenger, Mike, Ranjit Bharvirkar, Bill Golmeboski, Charles A. Goldman, Steven R. Schiller. *Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs.* Berkeley, CA: Ernest Orlando Berkeley National Laboratory (LBNL), Environmental Energy Technologies Division, 2010.

implementation staff are engaged; Energy Trust, NSTAR and NYSERDA obtain feedback from implementation staff before the evaluations are deemed final.

FINDINGS

Organization Background and Context of Efficiency Activities

The American Council for an Energy Efficient Economy defines an energy efficiency resource standard as a policy that sets annual energy efficiency targets over the long term.[14] Under this definition, all of the examined organizations operate under energy efficient resource standards. The organizations examined describe their energy efficiency goals in two distinct ways. Washington, Massachusetts and California require energy efficiency program administrators to obtain all cost-effective energy savings. In contrast, Pennsylvania and New York have established goals for each program administrator to ensure the state will meet mandated percentage reductions in energy use. In Washington, Massachusetts and Pennsylvania recent legislation has led to notable increases in energy efficiency goals and program budgets.

The majority of the organizations examined fund energy efficiency programs through a system benefit charge separate from utility rates. Pennsylvania utilities were the only examined organizations that consider efficiency program costs part of the utility's operating costs, to be recovered through rates. Avista and PSE fund efficiency programs through tariff riders.

Four organizations have some type of external advisory group that provides oversight of their evaluation activities. These groups are typically involved in the creation of EM&V plans and may review evaluation findings.

The comparison organizations face a range of regulatory requirements regarding evaluation of energy savings. These requirements, as well as characteristics of the organizations themselves, influence each organization's approach to program delivery and evaluation. We provide a brief summary of each organization below.

Avista

Avista, the smallest organization included in this analysis, is facing increased targets for energy conservation and renewable energy procurement following the passage of Initiative 937 (I-937) by Washington voters in 2006. I-937 directed large utilities in Washington to establish plans for acquiring all cost-effective energy conservation and meet annual goals for cost-effective resource acquisition. The requirements of I-937 took effect in 2010.

[14] Source: ACEEE, "Energy Efficiency Resource Standards (EERS)" *http://www.aceee.org/topics/eers*.

Avista has offered energy efficiency programs since 1978 and since 1995 has funded efficiency through a system benefits charge tariff rider. However, I-937 is expected to increase the amount of resources Avista devotes to program evaluation and alter its evaluation activities. Until 2008, Avista's efficiency programs operated under a *prudence standard* based on total resource cost. In 2009, Avista's regulatory structure shifted toward requiring more detailed impact evaluation, process evaluation and market effects analysis.

Avista has an External Energy Efficiency Board (the *Triple E* board) which provides advisory oversight to the utility's conservation activities, including evaluation. The Triple E board, through a collaborative process, produced a high-level EM&V Framework document detailing Avista's approach to evaluation. This document was filed on September 1, 2010 with the Washington Utilities and Transportation Commission (WUTC). The Triple E is also focusing on the utility's annual EM&V plans, and may review a forthcoming Technical Resource Manual. Avista also works with other regional organizations in conducting evaluation, including the Regional Technical Forum and the Northwest Energy Efficiency Alliance.

Energy Trust of Oregon

Energy Trust of Oregon is a non-profit organization that, under contract to the Oregon Public Utilities Commission (OPUC), provides programs to promote energy efficiency and renewable energy for customers of Portland General Electric, Pacific Power, Northwest Natural, and Cascade Natural Gas Corporation.[15] Launched in 2002, Energy Trust's funding comes through a range of legislated and administrative arrangements including:

- A portion of a 3% electric efficiency and renewable surcharge established under SB1149- 1.701% for efficiency and .513% for renewable energy
- Additional electric funding set at levels to meet Integrated Resource Planning efficiency goals.

Funding for natural gas efficiency programs is based on decoupling settlements with the OPUC, set at levels to meet those utilities' IRP goals. Energy Trust reports to the OPUC and the utilities that provide funding.

The OPUC does not directly regulate Energy Trust. Instead, Energy Trust operates under a grant agreement and contracts with the PUC and participating utilities. Energy Trust's contract with the PUC specifies minimum performance standards the organization must meet, and allows the PUC to send a notice of inquiry and ultimately cancel Energy Trust's contract if the organization fails to meet the standards specified in the contract. The participating utilities also set goals for

[15] Energy Trust of Oregon is also offering programs on a pilot basis for Northwest Natural Gas in Clark County, Washington.

Energy Trust's performance and incorporate savings obtained through Energy Trust as part of their IRP process. Energy Trust's board of directors provides the organization's direct oversight.

An evaluation committee of Energy Trust's board of directors reviews draft evaluation reports and provides comments before the reports are finalized. Two independent, volunteer expert advisors, who are well-known and credible individuals in the evaluation field, support the evaluation committee in these tasks.

NSTAR

The 2008 Massachusetts' Green Communities Act raised the energy savings goals of NSTAR and other efficiency program administrators in Massachusetts and spurred the implementation of new and expanded efficiency programs. The Act requires energy efficiency program administrators to collaborate to develop a statewide plan to achieve all cost effective energy efficiency. The Act created a per-kilowatt hour charge to fund energy efficiency and requires utilities to file three-year energy efficiency plans.

The Green Communities Act created an Energy Efficiency Advisory Council that plays a role in designing and approving energy efficiency programs. Program administrators are also required to collaborate with this Council to create a uniform, statewide EM&V plan. Efficiency program administrators in Massachusetts must submit evaluation plans with their three-year efficiency program plans and are required to include evaluation findings in their annual reports to regulators.

NYSERDA

The New York State Energy Research and Development Authority (NYSERDA) is a public benefit corporation that operates energy efficiency programs and conducts research into energy supply, energy-related environmental issues and research and development efforts related to energy.

Since 1998, a substantial portion of funding for NYSERDA's energy efficiency activities has come from a system benefits charge paid by New York ratepayers. New York's Energy Efficiency Portfolio Standard (EEPS), which took effect in 2008, increased NYSERDA's energy efficiency goals and provided the organization with funding to expand its energy efficiency offerings. NYSERDA also offers programs funded by the Renewable Portfolio Standard, the Regional Greenhouse Gas Initiative (RGGI) and the American Recovery and Reinvestment Act (ARRA).

A 13-member board oversees NYSERDA's activities. The board includes the commissioners of the New York Department of Transportation and Department of Environmental Conservation and the chairs of the Public Service Commission and the New York Power Authority. The Governor of New York appoints other board members.

Pennsylvania PUC

Numerical energy savings and demand reduction targets for each of Pennsylvania's seven investor-owned utilities with more than 100,000 customers were established in a 2008 state law, Act 129 that established an energy efficiency portfolio standard. Utilities that fail to meet their targets face penalties. Act 129 also requires each of the utilities to contract with an independent evaluator who will conduct that utility's evaluation activities and report the utility's energy efficiency and demand response accomplishments to the Pennsylvania Public Utilities Commission (PUC). The PUC has also hired a third-party statewide evaluator (SWE) to oversee and audit the evaluation activities managed by the utilities. While the SWE will not conduct evaluations of its own, it is expected to spot check and review records from the impact and process evaluations completed by each utility's evaluation contractor and verify a sample of measure installations.

The SWE will also work to develop evaluation protocols for custom measures expected to compliment Pennsylvania's existing technical resource manual, which provides energy savings values for deemed measures. In addition, the SWE will conduct a market potential survey to guide Pennsylvania's energy efficiency activities beyond 2013, defining the potential to achieve additional savings and recommending budget levels necessary to meet those savings goals. The *Audit Plan and Evaluation Framework for Pennsylvania Act 129 Energy Efficiency and Conservation Programs* (Audit Plan), prepared by the SWE, provides additional details about the role of the SWE.

San Diego Gas & Electric (SDG&E)

California's investor-owned utilities have offered energy efficiency programs to California ratepayers for many years. Since 1996 these programs have been primarily funded through a system benefits charge included in ratepayers' energy bills. In order to achieve all cost effective energy savings, the California Public Utilities Commission (CPUC) sets annual and cumulative goals for energy savings in each utility service territory, which it periodically updates. Since 2005, the CPUC has set these goals at the portfolio level in order to allow utilities to pursue pilot programs or innovative measures for which the energy savings may be difficult to predict. The CPUC approves utilities' energy efficiency program plans. The CPUC also requires that a set percentage of system benefit charge funds go to evaluation.

Program evaluation duties in California are divided between regulators and utilities. Since the beginning of the 2006-2008 program cycle, the staff of the CPUC's Energy Division and the California Energy Commission (the *Joint Staff*) has managed impact and market effects evaluation, leaving the utilities to oversee process evaluation. The CPUC must also approve process evaluation plans. In 2009 the CPUC expanded its role in process evaluation, and will oversee process evaluation if it determines it necessary to do so.

Organizational Structure

Contacts at each of the comparison organizations described how evaluation fits within their organization and how evaluation staff interact with program implementation staff. In interviews, our discussions of organizational structure focused on two primary topics. First, we sought to understand the reporting path and organizational chart for evaluation. Second, we sought to understand the responsibilities of the evaluation department and the scope of expectations for evaluation leadership.

We created simplified organizational charts for each of the comparison organizations with program administration responsibilities (excluding the Pennsylvania PUC). These diagrams are presented in Appendix A. In comparing the organizational structure around evaluation at each of these organizations, we identified four primary themes.

In all cases, energy efficiency programs and evaluation were housed together in the segment of the organization focused on delivering energy efficiency. Thus, at each of the four investor-owned utilities, evaluation and efficiency programs are housed under a director charged with providing a range of services to customers. These directors may have titles that reflect this broader responsibility: for example "Vice President for Customer Care" or "Director of Customer Programs" but not always. At PSE, this person is "Vice President Energy Efficiency Services," and Avista calls this person "Vice President Sustainable Energy Solutions." In all cases, however, the person at this level is responsible for more than program implementation and evaluation. They are charged with planning, marketing, market research and, in many cases, regulatory reports associated with integrated resource planning and energy savings estimates.

At the two non-utility program administrators, Energy Trust and NYSERDA, this director tends to be at or near the top of the organization chart, since energy efficiency is the primary focus of the organization. Thus, the Executive Director at Energy Trust oversees all functions associated with energy efficiency, as does the President of NYSERDA. However, both of these organizations separate program implementation from planning and evaluation functions at the organizational level immediately below them.

While these functions are housed in the same overall segment of the organization, ***it is common for evaluation, planning and market research to be organizationally separated from program implementation***—managed by leadership that exist at the same level within the organization. Among the organizations that implement programs, all but PSE have the evaluation and implementation staff report to different managers at the same level. At PSE, the Director of Market Strategies and Director of Customer Energy Management are at the same organizational level, but the evaluation function reports to the same director as implementation.

Evaluation is rarely separated from planning and market research. PSE is unique in separating strategic planning and market research from evaluation. Each of the comparison program administrators grouped evaluation with the market research and planning staff. This is not to say that evaluation staff also necessarily conducted planning and market research activities—rather,

that evaluation staff were housed under a director also responsible for broader energy analysis and planning activities.

Finally, ***concerns about independence*** and a desire for evaluation and planning staff to be able and willing to make hard calls about programs is a common reason for establishing separate reporting paths for evaluation and program implementation staff. Cooperative, communicative relationships between evaluation staff and program implementation were considered important, if not vital, by many of the contacts we interviewed. However, these relationships are expected to exist within an expectation of independence. Contacts at Avista report that consideration was given to moving the evaluation staff to different vice president entirely, but that costs and organizational considerations associated with this option caused it to be rejected.

Evaluation and energy analysis activities have the potential to support program effectiveness and simultaneously provide organizations with some assurance that the existing programs or measure mix will lead to goal attainment at a portfolio level. If data indicate that this may not be the case, evaluation staff are expected to communicate these findings to planning and program staff. In addition, evaluation staff members at the majority of the examined organizations work closely with resource planning staff to ensure that decisions related to resource allocation consider evaluation findings. At Energy Trust, NSTAR and NYSERDA, a single manager oversees both evaluation and planning activities to facilitate this type of close cooperation. The business planning process and high-level budgeting are included in the roles of Avista's evaluation staff members.

Role of Evaluation in Program Planning, Implementation and management

Evaluation Planning

Organizations may plan and conduct evaluation at a variety of levels. Evaluations focused at the measure level provide estimates of the energy savings attributable to individual measures or end uses. Programs are a means to deliver one or more measures to a market segment, and evaluations focused at the program level provide estimates of the energy savings resulting from all of the measures a program offers and often include process or market evaluation components to improve the efficiency and effectiveness of program delivery and implementation. Evaluation focused at the sector level examines results attributable to all of the programs serving a particular market sector, while evaluation focused at the portfolio level investigates results of all the programs an administrator offers. In contrast to the impact focus of most measure specific evaluation efforts, program level, sector level and portfolio level evaluations commonly address a variety of evaluation issues including process, market, impacts, theory and logic questions, and needs assessments.

Four of the seven organizations reviewed here plan evaluation at the program or measure level and later aggregate evaluation budgets and findings to the portfolio level. Two of these organizations noted that planning at the program level allows for evaluations to better account for the unique aspects of each program and provide more detailed findings. According to

Pennsylvania PUC staff, planning evaluation at the program level allows utilities to account for differences between customer groups and variation in the types of measures that each program supports. Similarly, Energy Trust may focus an evaluation on individual measures, program components, or other issues important to program success. Avista plans evaluation at the program level and later aggregates results in order to meet the different regulatory requirements in Idaho (which requires program-level reporting) and Washington (which allows portfolio-level reporting). PSE plans evaluation at the measure level to a greater extent than the other organizations examined, although PSE may group similar measures targeting a single customer group for evaluation.

SDG&E and NYSERDA approach evaluation planning a bit differently. In 2010, SDG&E grouped its programs by sector and hired evaluation contractors to conduct process evaluations at the sector level. From 1999 to 2007, NYSERDA as the 2007 Energy Public Benefits Program Evaluation Plan states hired contractors to carry out evaluation across the full portfolio of programs. By planning evaluation at the portfolio level, NYSERDA sought to better integrate evaluation results, achieve greater consistency in evaluation approach, allow for better coordination among evaluators, and provide administrative efficiencies. Since 2007, and an increase of funds allocated to evaluation from 2% to 5%, NYSERDA has increased its focus on evaluation planning at the program level, by developing detailed, multi-year evaluation plans for each program that the Department of Public Service reviews and posts on its website.

In planning evaluation, contacts from Energy Trust, NYSERDA, the Pennsylvania PUC and SDG&E emphasized the importance of providing process evaluation findings quickly. Contacts cited two reasons for this emphasis on the timeliness of evaluations. First, if findings are not timely, a program's environment or implementation practices may change, making findings irrelevant. Second, when findings are available quickly, program staff can incorporate those findings into program implementation before the end of the program cycle.

To ensure that program staff receive evaluation findings quickly enough that those findings remain relevant, contacts at Energy Trust and the text of Pennsylvania's Audit Plan urge providing program staff with evaluation findings even before an evaluation report is final. In order to make evaluation findings available in this way, Energy Trust conducts brief surveys of a sample of program participants each month to gather information on process issues and market effects. While the results of these surveys inform larger evaluation efforts, incremental results are available to program staff quickly through a tracking dataset.

In order to ensure that programs have time to adapt to evaluation findings, NYSERDA, SDG&E and the Pennsylvania PUC devote resources to process evaluation early in the program cycle. Pennsylvania's Audit Plan specifies that programs have typically stabilized enough for process evaluations to take place approximately six months into the program cycle. Similarly, SDG&E reported it initiates process evaluations six months into the program cycle with the goal of completing the evaluation half way through the three-year cycle.

Evaluation Budget

As noted in Table 2, the majority of the examined organizations devote between three and five percent of their energy efficiency program budget to evaluation. This finding is consistent with the evaluation budget levels listed in LBNL's *Review of Evaluation, Measurement and Verification Approaches*, which found that the majority of the states examined (8 of 15), devote between two and five percent of their program budgets to evaluation. PSE is the exception in this regard, with only one percent of its program budget devoted to evaluation in 2010. Three of the organizations examined operate under regulatory requirements that specify the portion of program costs devoted to evaluation. Because PSE and Avista allocate costs differently, the two percentages may not be appropriate comparisons—Avista includes program-level verification activities in the utility's 2010 evaluation budget, while PSE is only including independent evaluation expenditures. PSE and Avista are currently working with stakeholders and Washington regulators to establish evaluation spending targets.

Table 2: Annual Efficiency Program Spending and Evaluation Budgets

Organization	Total 2010 Efficiency Budget	2010 Evaluation Budget	Evaluation Spending as % of Efficiency Program Budget	
			Target	2010
Avista	$25,273,957	$1,000,000	3-6%	4.0%
Energy Trust	$112,310,279	$4,585,404	None Specified	4.1%
NSTAR	$121,716,273	$4,500,000	4%	3.7%
NYSERDA	$389,389,714[i]	$22,411,031[i]	5%	5.8%
Pennsylvania	$231,258,640[ii]	Not Known	None Specified	Not Known
SDG&E	$74,647,901	$2,965,333	4%	4.0%
PSE	$98,715,000	$947,500	1-3%	1.0%

[i] Includes SBC, EEPS, and RGGI funding. SBC funding figures from State of New York Public Service Commission *Case 05-M-0090 – In the Matter of the System Benefits Charge III.* Appendix C. EEPS funding figures from State of New York Public Service Commission *Case 07-M-054 – Proceeding on Motion of the Commission Regarding an Energy Efficiency Portfolio Standard*. Table 15. RGGI funding figures from NYSERDA, *Operating Plan for Investments in New York under the CO_2 Budget Trading Program and the CO_2 Allowance Auction Program,* June 21, 2010. Table 3-2.

[ii] Act 129 caps utility efficiency spending at 2% of revenues as of December 31, 2006. This figure reflects 2% of revenues of the seven utilities subject to Act 129 as reported in the 2006 EIA-861 database.

Contacts at three organizations reported tracking evaluation expenses as part of program costs, while two organizations reported tracking evaluation costs separately from program costs or administrative costs (Table 3). Avista tracks evaluation costs different ways for different purposes—considering evaluation a program cost for cost effectiveness calculations but a stand-alone expense to comply with regulatory requirements that specify evaluation spending levels. Avista also uses the proportion of energy efficiency program costs paid in incentives as a metric to determine the ultimate customer benefit an energy efficiency budget provides. In those

calculations, Avista includes evaluation with other administrative expenses as 'non-incentive' costs.

Table 3: How Evaluation Costs are Applied and Incorporated

Organization	Categorization of Evaluation Costs	Level at Which Evaluation Costs are Included in Cost Effectiveness Tests	
		Program	Portfolio
Avista	Varies by Application	X	
Energy Trust	Program	X	
NSTAR	Program	X	
NYSERDA	Stand-Alone Expense	X	
Pennsylvania	Program		X
SDG&E	Stand-Alone Expense		X
PSE	Stand-Alone Expense		X

We also sought to understand whether or not evaluation costs were included in cost effectiveness calculations for individual programs. California and Pennsylvania do not require individual efficiency programs to meet cost effectiveness standards, although each utility's overall portfolio must pass cost effectiveness tests. In both states, this focus on portfolio-level cost effectiveness is designed to allow utilities to offer programs targeting customers with limited incomes, pilot programs, education programs and other types of programs that might not provide cost-effective savings. In some cases, regulators require utilities to offer these types of programs.

As a result of this focus on cost effectiveness at the portfolio level, SDG&E evaluation staff report the utility does not examine cost effectiveness of individual programs unless it becomes necessary in order to cut costs to achieve a cost effective portfolio.

Allocation of Evaluation Resources

By Program Type

Among contacts that reported allocating evaluation resources by sector, all but PSE reported devoting more resources to the commercial and industrial sectors. At NYSERDA and NSTAR, this allocation reflects the fact that commercial and industrial programs provide the largest amount of energy savings. In addition, these programs tend to rely on a mixture of prescriptive and custom elements, increasing the complexity and cost of evaluation. The Pennsylvania PUC expects that utilities in Pennsylvania will devote the largest portion of their evaluation funding to commercial and industrial programs for similar reasons. At Avista, contacts explained that programs in which a substantial amount of savings are attributed to custom programs require more evaluation resources.

PSE differs from NSTAR and Pennsylvania utilities in the sense that the majority of PSE's planned evaluations, and its projected evaluation costs, focus on the residential sector. PSE staff members cite two reasons for the organization's focus on evaluation of residential programs. First, PSE has developed new programs and pilot programs targeting the residential sector more quickly than it has developed new programs for commercial and industrial customers and evaluation resources are often allocated to new programs. Second, like NSTAR, Avista and Pennsylvania PUC staff, PSE staff noted that, because of their more frequent use of custom measures, commercial and industrial programs require more complex evaluation efforts than residential programs drawing primarily on prescriptive measures. According to evaluation staff, PSE does not have the capability to effectively and efficiently evaluate large numbers of custom projects.

Energy Trust and SDG&E did not specify whether their allocation of evaluation resources varies by program type.

By Evaluation Type

All of the organizations examined for this research devote the largest portion of their evaluation budgets to impact evaluation. Contacts from NYSERDA, SDG&E and PSE elaborated that their organizations spend approximately 65-70% of their evaluation budget on impact evaluation. Findings of the LBNL *Review of Evaluation, Measurement and Verification Approaches* suggest that this allocation is typical. All but one of the states examined in the LBNL study reported devoting the majority of their evaluation resources to impact evaluation, and the majority (6 of 11 states providing data) reported that between 60% and 80% of their evaluation budgets go to impact evaluation. Explaining this focus on impact evaluation, contacts from NYSERDA and the Pennsylvania PUC cited the need to comply with mandated savings goals and meet statewide energy-use reduction targets. The CPUC sets SDG&E's evaluation budget and withholds a portion of that budget to conduct impact evaluation.

While all of the examined organizations devote the largest portion of their evaluation funding to impact evaluation, contacts emphasized that their organizations nonetheless value process and market studies. Avista staff noted that third-party contractors will conduct the organization's impact evaluation efforts while internal staff will carry out process evaluation at very little incremental cost to the company. As a result, Avista staff stated that their organization values process evaluation more highly than their evaluation budget would indicate. NSTAR staff anticipates that funding for process evaluation will increase as the organization implements pilots and new programs and tries new approaches in response to the Green Communities Act.

Prioritization of Evaluation

The extent to which a program contributes to the overall portfolio of savings and the uncertainty surrounding savings estimates are primary factors that drive evaluation priorities. Contacts cited a variety of program characteristics that can contribute to the uncertainty of savings estimates, including:

- Lack of information related to measure performance, energy savings, and persistence.
- The potential variability of free ridership rates as market conditions change.
- Uncertain measure operating conditions, which are in turn influenced by user behavior, weather, and the effectiveness of the installation.
- Factors related to program management like the effectiveness of messaging, the need for quality control and program requirements.
- The proximity of program measures to cost effectiveness targets, with measures that achieve cost effectiveness ratios closer to one introducing greater uncertainty to program level cost effectiveness estimates.
- The length of time since a previous evaluation and the extent of program change in that time.

In addition to program characteristics, contacts noted that evaluation methodologies may also result in uncertain savings estimates. One contact elaborated that, in evaluation planning, an effective survey approach and sufficient coefficient of variation are assumed, and incorrect assumptions can result in gathering sub-optimal data or failing to achieve targeted confidence and precisions levels. In addition, instrumentation error and sampling error may lead to uncertainty.

Contacts seek to design evaluation studies in ways that will mitigate this uncertainty. Guidelines in California and Pennsylvania ask evaluators to quantify risks that they cannot control. NYSERDA and California's Joint Staff have used these quantified risks to conduct risk analyses that help inform their prioritization of evaluation resources. A contact from NYSERDA also reported closely monitoring the progress of evaluations in order to quickly address unanticipated challenges.

Contacts also reported considering the cost of evaluation and the potential to mitigate risk when prioritizing evaluation. According to one evaluation staff member, "You might have high uncertainty, but there may not be much you can do without throwing a lot of money at it that you don't have."

In addition to risk to the overall savings portfolio, the examined organizations consider a variety of factors in setting evaluation priorities. These factors include:

- Information requests from program staff or other stakeholders and the potential for evaluation results to benefit program implementation.
- Evaluation studies necessary to gather data required for regulators or other types of oversight.
- The potential to leverage other work or expand the evaluation results produced by an external organization (such as the Northwest Energy Efficiency Alliance).

➡ The future of the program, including whether it is expected to continue or if implementation is likely to change substantially.

Expectations of Evaluation Products

Protocols

As the expectations and legal requirements tethered to energy efficiency programs expand, protocol documents and technical reference guides are becoming increasingly common. These documents are expected to improve the certainty of estimated energy savings and energy efficiency acquisition and ensure that the rapidly expanding budgets for efficiency programs are well spent. Protocols also enable comparison of energy efficiency results across utilities subject to the protocols. As is visible in Table 4, among the comparison organizations, only those in the Pacific Northwest are operating without an existing evaluation protocols document. As regulators in Washington finalize the I-937 Conditions for the Washington utilities affected, even this is expected to change.

New York's evaluation guidelines list the types of data an evaluation would likely gather and references the National Action Plan for Energy Efficiency's Model Energy Efficiency Program Impact Evaluation Guide as a guideline for evaluation methodology.[16] New York also has a series of Technical Manuals that specify methods for estimating energy savings.

Table 4: Summary of Evaluation Factors

Organization	Protocols Established	Confidence & Precision Level Required	Confidence & Precision Sought	Frequency Specified
PSE	In process	No	90/10	No
Avista	In process	No	90/10	No
NStar	Yes	No	90/10	Through stakeholder process
SDG&E	Yes	Yes, with caveats	Varies	Through Joint Staff
NYSERDA	Yes	Yes	90/10	No
Energy Trust	No	No	Varies	No
PA PUC	Yes	Yes	Varies	No

[16] The National Action Plan for Energy Efficiency was a collaborative effort involving gas and electric utilities, utility regulators and other organizations that sought to bring about a national commitment to energy efficiency.

Sampling Confidence and Precision

Sampling confidence and precision are measures of the reliability of the sampled data, that is, they are indicators of how well the sample represents the population and how confident the researcher is that the estimated values represent true values for a given population. The level of confidence and precision expected from a given evaluation is a function of the sample size, the degree of variability in a given measurement (for example in the number of responses to a question or in estimates of energy use), and, to a lesser degree, in the size of the sample as a proportion of the population. The presence of protocols does not necessarily equal firm expectations for sampling confidence and precision, however 90/10 confidence and precision emerged as a common goal.[17]

Since confidence and precision are affected by the variability in measurement as well as the sample size, two measurements from a given sample (e.g., the responses from two questions in a survey) will not necessarily have the same confidence and precision levels. Evaluators typically estimate the sample needed to achieve a given confidence and precision level under a particular assumption about the level of variability.

The level of confidence and precision ultimately obtained can be driven by budget considerations as well as the measure characteristics. Obtaining 90/10 confidence and precision on a measure that contributes very little to the overall portfolio of energy savings or a single question in a survey may be cost prohibitive for the overall benefit to the organization. All of the contacts interviewed as part of this effort mentioned budget constraints and the value of effectively designed evaluation when discussing technical expectations for evaluation. Contacts also noted that program staff may raise concerns when evaluation studies do not meet high levels of precision yet produce results that suggest lower realization rates than program staff expect.

Two contacts specifically described having to resolve conflicts that result from evaluation findings in which a high level of confidence and precision were either not feasible or not obtained. At PSE, program staff have requested a review of evaluation findings when high levels of confidence and precision were not obtained.

In California, evaluation protocols recognize that evaluation budgets will limit the size of the sample evaluators can draw and as a result limit the level of precision an evaluation can achieve and the extent to which evaluators can minimize bias. The Joint Staff allocates evaluation resources with the goal of reducing overall portfolio uncertainty and assigns levels of rigor to each evaluation accordingly. Consistent with this approach, the protocols do not penalize evaluators for failing to meet the confidence/precision levels the protocols suggest.

---

[17] The California Sampling Protocol however specifies that 90/10 confidence and precision is not necessarily sufficient when the mean and standard deviation is available and the coefficient of variation suggests a preferable sampling strategy.

In their effort to minimize risk to the statewide portfolio of energy savings, the Joint Staff oversees impact evaluations that generate average realization rates and savings values at the measure level. However, utility program administrators in California have criticized this approach, stating that statewide average values do not accurately reflect the unique conditions in which their programs operate. Contacts at SDG&E reported frustration among utility staff when impact evaluations failed to achieve sufficient confidence and precision levels, and thus provided less-than-reliable estimates of energy savings resulting from the 2006-2008 programs.

Sampling confidence and precision affect the overall level of certainty afforded to evaluation results. Process evaluations, frequently used to understand the strengths and weaknesses in program implementation, may not be expected to achieve 90/10 in every case. Impact evaluations, particularly when shareholder incentives or resource adequacy plans are affected by the outcome, are typically expected to achieve high levels of confidence and precision. At NYSERDA, Energy Trust, and the Pennsylvania PUC, the appropriate or practically achievable levels of confidence and precision are established by working directly with expert evaluators, who serve in an advisory role like Energy Trust's board evaluation committee and NYSERDA's System Benefits Charge Advisory Group. These expert evaluators consider the overall level of risk to the success of the portfolio in setting targeted confidence and precision levels. In Pennsylvania, the statewide evaluator guides the PUC and provides technical review of evaluation products. Contacts at Avista believe this could be useful in Washington State to avoid the controversy that results from review by non-technical stakeholders, or those without evaluation background.

Frequency

As displayed in Table 5, only SDG&E operates with a specific expectation of evaluation frequency, noting that the Joint Staff expects every program to receive a process and impact evaluation at some point in each three-year program cycle, however comments for each of the organizations suggest evaluations occur for every program within each program cycle.

Table 5: Estimated Frequency

Organization	Evaluation Frequency
PSE	Every four to five years; more frequent if needed
Avista	Annual billing analysis; other evaluation as necessary
NStar	Impact evaluations at least every two years. Other evaluation needs determined in cooperation with stakeholders.
SDG&E	Determined by Joint Staff. Resource programs must receive process and impact during each 3-year program cycle.
NYSERDA	Driven by program characteristics.
Energy Trust	Driven by program characteristics, but annual is common.

Organization	Evaluation Frequency
PA PUC	Process: driven by program characteristics. Impact: some form of continual assessment expected.

As previously noted, SDG&E divided the process evaluations of its 2006-2008 programs by customer sector, but these evaluations provided findings related to each program within a sector. California regulators also form groups of similar programs or programs targeting similar sectors for impact evaluation, although the Joint Staff allocates evaluation resources for impact evaluation at the program and measure level.

Evaluation staff members at PSE reported that programs typically receive evaluations every four to five years, however, if necessary a program might receive some level of evaluation attention more frequently, if not annually. Pilot programs (which, at PSE typically target single measures), those in which market conditions or program design have changed, or programs without stable savings assumptions or realization rates would be candidates for more frequent evaluation. Avista files an annual evaluation plan for its DSM activities that includes a bill verification analysis to identify any changes in customer usage attributable to DSM programs. NSTAR determines the frequency of program evaluation as part of a broader stakeholder process that sets evaluation priorities. Like PSE, contacts at NSTAR report that the frequency of program evaluation, especially process evaluations, is driven by program design or market changes and the stability of a program's savings attainment. Typically, though, a program would receive an impact evaluation at least every two years, for C&I programs impact work might focus on a specific end use rather than the program as a whole.

New York's guidelines do not include specific requirements regarding evaluation frequency. Instead, the guidelines specify characteristics that may justify more or less frequent evaluation for individual programs. In New York, evaluation frequency also reflects the pace of a program, both in projects or in spending. For example, programs delivering large amounts of energy savings through a relatively small number of large projects that occur over a long timeframe may receive less frequent evaluation. The evaluation guidelines suggest that program administrators focus on process evaluation early in a program's history since the findings may identify opportunities to improve program performance and because new programs may not have completed enough projects for an impact evaluation to be meaningful.

Most of Energy Trust's major programs receive impact and process evaluations every year, although impact evaluations may occur less frequently if a program's realization rate remains relatively constant. Similarly, impact evaluations may focus less attention on measures or program elements for which savings are relatively well understood and instead investigate a particular measure or a new application that is creating uncertainty. The frequency of process evaluations depends to a greater extent on the characteristics of a program and its information needs. Process evaluations typically occur annually, but often focus on different issues from one year to the next. Energy Trust staff also consider evaluation activities that the Northwest Energy Efficiency Alliance undertakes, which may include an examination of regional or cooperatively implemented programs.

The Pennsylvania PUC is just beginning to manage the evaluation process established by Act 129. Contacts expect that monthly and quarterly reports will help them track program progress and allow for effective allocation of funding and staff resources. The ongoing evaluation and reporting process is designed to facilitate the statewide evaluator's (SWE) efforts to audit evaluation results. Process evaluations are expected to occur less frequently than impact evaluations, and are viewed as an important tool for assessing programs early in their life cycle. According to the Audit Plan, programs have sufficient stability after approximately six months of implementation to allow evaluators to accurately gauge the effectiveness of program processes. The plan also notes the benefits of feedback mechanisms that provide findings even before an evaluation report is final.

Overall Credibility

Regardless of the overall approach, all of our comparison contacts reported being satisfied with the credibility of the evaluation products they are responsible for. In California, SDG&E contacts are pleased with the quality of the process evaluations their organization manages. According to evaluation staff, SDG&E has been able to conduct process evaluations quickly enough that the findings are still relevant to program managers. California's centralized impact evaluation efforts tend to occupy a large portion of the third party evaluation contractors operating in the state, and in some cases this has affected the speed with which process evaluations could be completed. To overcome this, SDG&E has become more proactive in preparing for process evaluations, planning for evaluations and hiring consultants early in the program cycle. The CPUC manages impact evaluations in California, and SDG&E contacts noted that these evaluations would be more credible if they met stringent sampling confidence and precision thresholds.

NYSERDA's evaluation manager is satisfied with the credibility of the evaluation products her organization produces. According to the evaluation manager, NYSERDA's evaluation products gain credibility by: using competitively-selected third-party contractors; operating with an internal Energy Analysis group made up of experienced staff with a reputation for providing credible and objective analysis; and separating Energy Analysis staff from program implementation staff. NYSERDA also engages stakeholders in the evaluation process by incorporating advisory groups in evaluation planning and providing evaluation plans and other documents to the DPS and its evaluation contractor for review.

At Energy Trust, contacts believe their evaluation products are perceived as credible because the review process for evaluation products has helped the organization produce high-quality evaluation results that have leant credibility to the organization as a whole. In Pennsylvania, PUC staff are pleased with the credibility of the evaluation products produced under the new requirements so far, stating that the SWE's oversight increases staff members' confidence in evaluation findings. As of August 2010, utilities have submitted only preliminary evaluation reports. However, according to PUC staff, the SWE has identified issues related to sampling and related to realization rates for deemed measures.

Use of Evaluation Results

Communication with Program Staff

Contacts from the majority of comparison organizations reported that program implementation staff are involved in the process of prioritizing evaluation and setting the scope of evaluation studies, particularly in the case of process evaluation. The extent of program staff involvement in the development of evaluation efforts varied somewhat across the organizations. At NSTAR, evaluation staff work to inform program staff about evaluation activities through periodic meetings that cover program areas slated for evaluation and share information related to evaluation milestones. At Energy Trust, SDG&E, and NYSERDA, program staff take an active role in shaping evaluation studies. Evaluation staff may work with program staff to design evaluations that address information needs identified by program staff. Program staff may also be involved in developing requests for proposals soliciting third-party evaluation contractors.

Contacts from NSTAR and NYSERDA reported that third-party evaluators present initial findings to both evaluation staff members and program staff members. These contacts noted that such presentations allow program staff to ask questions or raise issues related to the evaluation. Since these presentations take place when evaluation findings are in a 'draft final' phase, evaluators have an opportunity to address program staff comments before submitting final evaluation reports. Energy Trust noted that providing draft evaluation findings to program staff members, even incrementally as report chapters become available, helps to ensure that staff receive evaluation findings quickly enough that recommendations remain relevant to program implementation.

Once evaluation reports are final, Energy Trust, NSTAR, NYSERDA, and PSE have a process by which program staff respond to evaluation findings, specifying ways they will adapt program implementation or justifying decisions not to implement recommendations. This process typically occurs through discussions between evaluation staff, implementation staff, and management. Energy Trust, NSTAR, and NYSERDA include information on program changes stemming from evaluation findings in regulatory filings or other reports to stakeholders or evaluators.

Avista differed from other organizations in how impact evaluation findings are communicated to program staff. In order to ensure the rigor and independence of evaluation results, implementation staff members have little opportunity to comment on impact evaluation findings before the report is final. Because Avista conducts process evaluations in-house, program staff will have more involvement in process evaluation studies as they take place. Avista expects program staff to adapt programs in response to evaluation findings, potentially discontinuing programs found to be underperforming or emphasizing programs found to provide more savings than anticipated.

Avista's approach contrasts with that of the other organizations, for instance the NYSERDA evaluation manager noted that program implementation staff are allowed to determine whether or

not they will adapt programs in response to particular evaluation findings. While NYSERDA expects staff to ensure that programs operate cost effectively, program staff may opt not to implement evaluation recommendations based on their assessment of other factors influencing the program and its environment. Program staff must justify decisions not to implement evaluation recommendations in a memo to management.

Resolution of Conflict Surrounding Evaluation Findings

Frequency of Conflict

The organizations examined for this research differed in their reports regarding the frequency with which program implementation staff have challenged evaluation findings. Contacts from Energy Trust, NYSERDA, PSE and SDG&E reported that it was common for program implementation staff to challenge evaluation findings that provide an unexpected result or characterize results in an unanticipated way. One contact noted that these challenges range from requests for minor clarification to more serious objections to evaluation methods.

While NYSERDA staff reported that process evaluation findings are more likely to receive challenges from program staff because they are typically more subject to interpretation than impact evaluation findings, evaluation staff from Energy Trust and SDG&E stated that impact evaluation findings receive the most serious objections. Contacts at Energy Trust report that program staff have at times argued that findings related to free ridership do not reflect their experience in the field and have objected to the methodology used to measure free ridership. SDG&E staff stated that evaluations conducted on a statewide-level, as impact evaluations are in California, cannot account for variables like the unique approach and management structure of each program. As a result, program staff in San Diego may not accept statewide impact evaluation results as accurate.

Evaluation staff at NSTAR reported that by maintaining regular contact with evaluation contractors, evaluation staff are able to inform program staff of unexpected evaluation findings before evaluation reports are finalized. As a result of this communication, evaluation findings rarely face challenges from program staff. In addition, NSTAR evaluation staff reported that most evaluations confirm savings achievements and market conditions consistent with program staff members' experience.

Consistent with its efforts to ensure the rigor and independence of evaluation findings, Avista typically does not seek revisions to evaluation reports when those reports provide unanticipated results or results that do not meet staff expectations. However, if Avista staff believe the results do not reflect a complete understanding of programs or projects, future projects may include revisiting an issue from a previous evaluation.

Conflict Resolution

Contacts at all of the comparison organizations reported working to resolve conflicts surrounding evaluation findings sufficiently for the findings to be finalized and used. Contacts from Energy Trust, NYSERDA and PSE described having an established process for overcoming challenges and finalizing evaluation results. These processes typically seek to address challenges while maintaining the integrity of evaluation research. Pennsylvania and California do not have formal processes for resolving challenges to statewide evaluation findings, beyond allowing utility staff to submit comments. Contacts from NSTAR and Avista reported that their evaluation findings rarely face challenges from program staff.

Energy Trust's process for finalizing evaluations and resolving challenges to evaluation findings draws on the oversight of the organization's board of directors. According to evaluation staff, this oversight helps to ensure that the primary motivation for any changes to evaluation findings resulting from staff comments is to maximize the integrity of the evaluation. The board's Evaluation Committee considers evaluation findings. Program staff are invited to attend meetings of the Evaluation Committee to provide comments and clarification related to evaluation results. According to evaluation staff, this process can lead to a relatively quick resolution of challenges when evaluation research is strong. However, if there are challenges to the evaluation's approach or interpretation of data and the evaluation committee is not satisfied with the evaluation methodology, Energy Trust may undertake additional research, which could take as long as six months to complete.

NYSERDA relies on the organizational separation between evaluation staff and program implementation staff as well as an evaluation contractor's interest in ensuring that the data presented in evaluation reports are accurate and unbiased to ensure the integrity of evaluation findings. According to evaluation staff, NYSERDA will allow program implementation staff to make comments on a few drafts of an evaluation report, but the evaluation contractor maintains control over the final contents of the report.

PSE's process for resolving conflict related to evaluation findings centers around the organization's Evaluation Response Reports (ERR). This process takes place after an evaluation report is finalized. Therefore, the ERR process does not have the potential to influence how data are presented or characterized in the evaluation report itself. Through the ERR process, evaluation staff work with program staff to determine how a program will respond to evaluation findings, although managers may be involved in decisions related to particularly controversial results. Evaluation staff report that this process has improved communication between program staff and evaluation staff and facilitated efforts to resolve conflict related to implementation of evaluation findings.

California and Pennsylvania do not have a defined process whereby program implementation staff can influence impact evaluation findings at a statewide level. As of August, 2010, Pennsylvania was beginning to generate evaluation findings under its new approach to energy efficiency. PUC staff anticipate that there will be an opportunity for utilities to respond to evaluation findings and recommendations by the statewide evaluator and that there may be an

opportunity for the evaluator to revise findings after consultation with utilities if evaluation data support doing so. According to staff, if the need arises, the PUC may define a formal process for this type of dialog.

The CPUC solicited utility staff members' comments on impact evaluation findings for the 2006-2008 program cycle, but the final impact evaluation reports did not address those comments to the extent SDG&E staff members had hoped. Beyond this process of providing comments, there is no formal process for SDG&E and other utilities to challenge the CPUC's impact evaluation findings. As a result, as of August 18, 2010, California utilities were still negotiating savings values and realization rates for 2006-2008 programs.

Stakeholder Use of Evaluation Results

Program Planning

In addition to adapting program implementation in response to evaluation findings, the comparison organizations use evaluation results in program planning and cost effectiveness calculations. As noted above, evaluation staff are typically closely involved in their organizations' planning activities, although PSE is an exception in this regard. At PSE resource planning is conducted by the Strategic Planning and Research group and program implementation staff carry out most program planning activities. In most organizations, contacts reported that evaluation findings contribute to program planning efforts in that they inform the assumptions that go into savings forecasts. Cost effectiveness considerations based on evaluation findings can also influence program budgets, potentially forcing program managers to cut costs.

Energy Trust has two defined processes that illustrate this use of evaluation findings in program planning. First, each February, Energy Trust produces a *True-Up* report, which applies evaluation findings to savings estimates and adjusts past program accomplishments to reflect new findings. Second, in August, Energy Trust's Planning and Evaluation Group uses evaluation findings to adjust the assumptions that go into the savings forecasts that ultimately inform the organization's annual budget.

The Pennsylvania PUC also considers evaluation findings as it reviews and approves each utility's annual program plans. According to PUC staff, in reviewing plans, staff members seek to ensure that each program is implementing changes consistent with identified best practices.

Rates and Incentives

Contacts from the majority of the organizations examined for this research (4 of 7) reported that evaluation findings have the potential to influence rates only to the extent that findings alter the amount of resources required to achieve energy savings goals. PSE and Pennsylvania utilities include the cost of energy efficiency programs in their operating costs, which are recovered in rates. The remaining organizations fund energy efficiency programs through a surcharge applied

to rates and staff of these organizations noted that changes in the level of resources required to achieve goals might justify an adjustment to the surcharge.

Organizations for which evaluation results might influence rates in other ways include Avista, SDG&E and NYSERDA. For Avista, the amount of fixed cost, which determines lost-margin recovery, is a function of verified energy savings and ultimately informs the decoupling rate adjustment. SDG&E conducts an analysis that includes the cost of running efficiency programs as well as the avoided cost of the conserved energy in determining rate impacts of its energy efficiency programs.

Only SDG&E is currently eligible for performance incentives based on its evaluated energy savings. Under Washington's I-937 PSE and Avista have the option to negotiate new incentive mechanisms. New York utilities can also receive incentives for documented energy savings, however as a public entity, NYSERDA is not eligible for these incentives. Utilities in Pennsylvania and Washington could face penalties if they fail to meet mandated energy savings targets.

Rather than being motivated by financial incentives, the non-utility program administrators examined (Energy Trust and NYSERDA) seek to meet contractual obligations and maintain funding. Were Energy Trust to consistently fall short of its goals, the Oregon Public Utilities Commission could send a notice of inquiry and ultimately cancel Energy Trust's contract. Similarly, the New York Public Service Commission could discontinue funding for under-performing NYSERDA programs.

Other Stakeholder Uses of Evaluation

NYSERDA is the only organization that currently has conducted an evaluation of its evaluations. Conducted twice so far, and planned for 2011, NYSERDA has examined the usefulness of the evaluations through a survey of program management and internal stakeholders. The results of this evaluation provided suggestions to the evaluators about how to improve their conclusions and recommendations and led to the development of the internal process NYSERDA uses for program staff to formally respond to the evaluation recommendations.

Role of Evaluation Contractors

Among the organizations compared here, only Avista conducts a majority of evaluation activities in-house. At PSE, NStar, SDG&E, NYSERDA, Energy Trust, and the Pennsylvania PUC, third-party evaluation contractors conduct most, if not all, of the evaluation work. At these organizations, the evaluation staff are primarily focused on contract management and typically work closely with both program staff and evaluation contractors to frame the scope of work, identify research questions to be addressed, and coordinate data requests.

Table 6: Approach to Evaluation Contractors

Organization	Portion Outsourced	Role of Staff
PSE	~90%	Contract managers; limited in-house evaluation; less in the future
Avista	Limited; primarily focused on impact evaluations of programs with high uncertainty	Extensive role in verification, analysis. Most process and market evaluations conducted in-house
NStar	~100%	Contract managers;
SDG&E	More than 90%	Contract managers; may conduct small process evaluations and basic impact analyses to inform savings assumptions
NYSERDA	~100%	Contract managers; actively engaged in research design discussions and data requests with contractors
Energy Trust of Oregon	~90%	Contract managers; will conduct small process evaluations and conducts billing analyses
PA PUC	~100%	Contract managers;

PSE, SDG&E and Energy Trust contract out most of their evaluation work, but will step in and conduct research projects in-house when appropriate. PSE contracts with third-party evaluators to conduct most of the organization's program evaluations, but staff will conduct limited process and impact evaluations. Third-party evaluation is particularly important when a project requires a skill set or level of expertise beyond that of the evaluation staff or when a project is simply too large. However, because in-house evaluations consume a great deal of staff time and the organization is expecting to increase the number and scope of evaluation work in the future, contacts at PSE expect that in-house evaluation will become impractical.

SDG&E and Energy Trust have a similar approach: relying on contractors for most evaluation work but retaining the ability to conduct energy savings analyses and small process evaluations. For SDG&E, the energy savings analyses would typically be limited to measures that had not yet been added to the statewide DEER database. Energy Trust conducts most billing analyses in-house. In California and at Energy Trust, the contracted evaluation model was selected to offer credibility of results. Contractors are engaged to review the work plans and evaluation products produced in-house at Energy Trust.

In the Northeast, NYSERDA, NSTAR and the Pennsylvania utilities expect that evaluation contractors will conduct all of the evaluation work. In these cases, evaluation staff are contract management experts and work closely with evaluation contractors to ensure that the research objectives meet the needs of the organization and the program staff. This is most visible at NYSERDA, where evaluation staff members work with contractors to design evaluation plans that meet applicable protocols; provide the data NYSERDA requires; review data collection instruments; and participate in methodological discussions. NYSERDA's evaluation staff is also expected to manage data requests, explain any nuances in the data, and facilitate conversations

with program implementation staff. The Pennsylvania PUC has a limited history with evaluation, but has established a formal review process that relies on a third-party evaluation contractor.

Avista reports the least extensive use of third-party contractors. At Avista, staff in the utility's operations group will conduct a sample of pre- and post-installation verifications and complete process tracking activities expected to inform process analyses. Evaluation staff, housed in the policy group, will conduct the analyses required to support decisions about cost effectiveness, will verify impact, process, and market evaluations conducted internally, and will oversee impact evaluations conducted by third-party contractors—typically engaged to review the utility's largest programs or those programs with the greatest uncertainty. Evaluation consultants are also engaged to review most internal analyses and evaluations and verify portfolio savings.

APPENDICES

APPENDIX A: ORGANIZATION CHARTS AND DESCRIPTIONS

APPENDIX B: EVALUATION UTILITY FACTORS

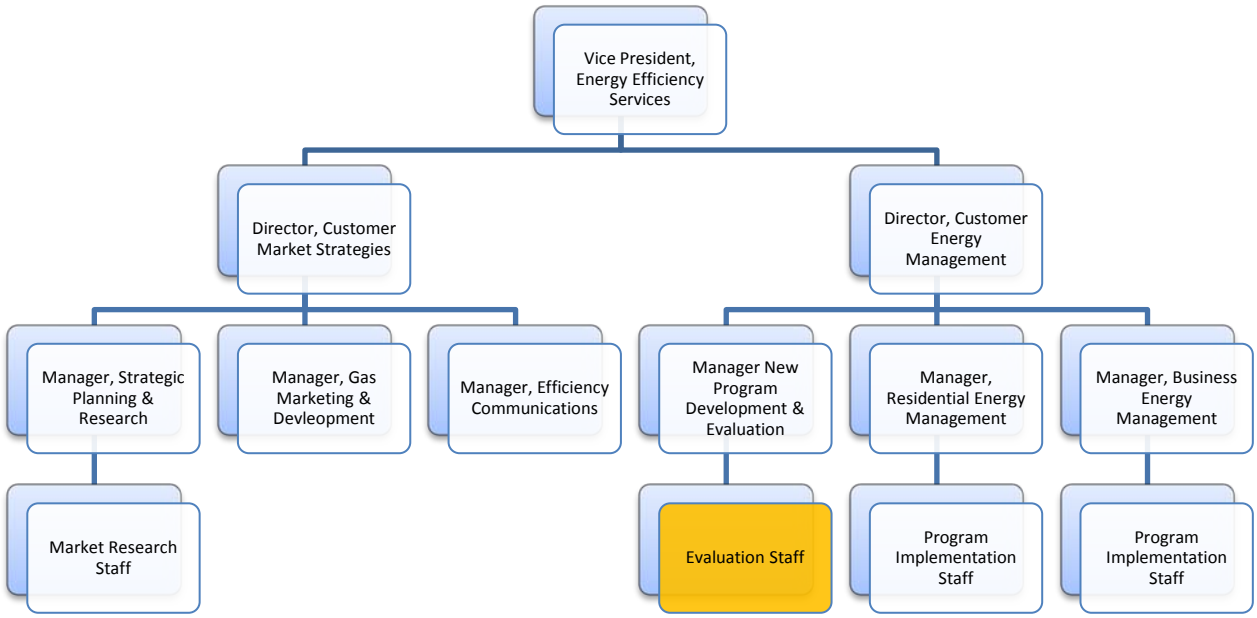
APPENDIX C: EVALUATION GLOSSARY

A ORGANIZATION CHARTS & DESCRIPTIONS

Puget Sound Energy

Puget Sound Energy is a Washington-based investor owned utility. At PSE, evaluation is the responsibility of the Manager of New Program Development and Evaluation. The evaluation group is responsible for program evaluation and cost effectiveness calculations, while the Manager of New Program Development also oversees PSE's demand response pilot programs and provides optimization support for existing programs. Other new program and pilot program development work is carried out by program implementation staff. The evaluation manager and the program implementation managers report to the Director of Customer Energy Management. Market characterization research is divided between the evaluation group and the Strategic Planning and Research group.

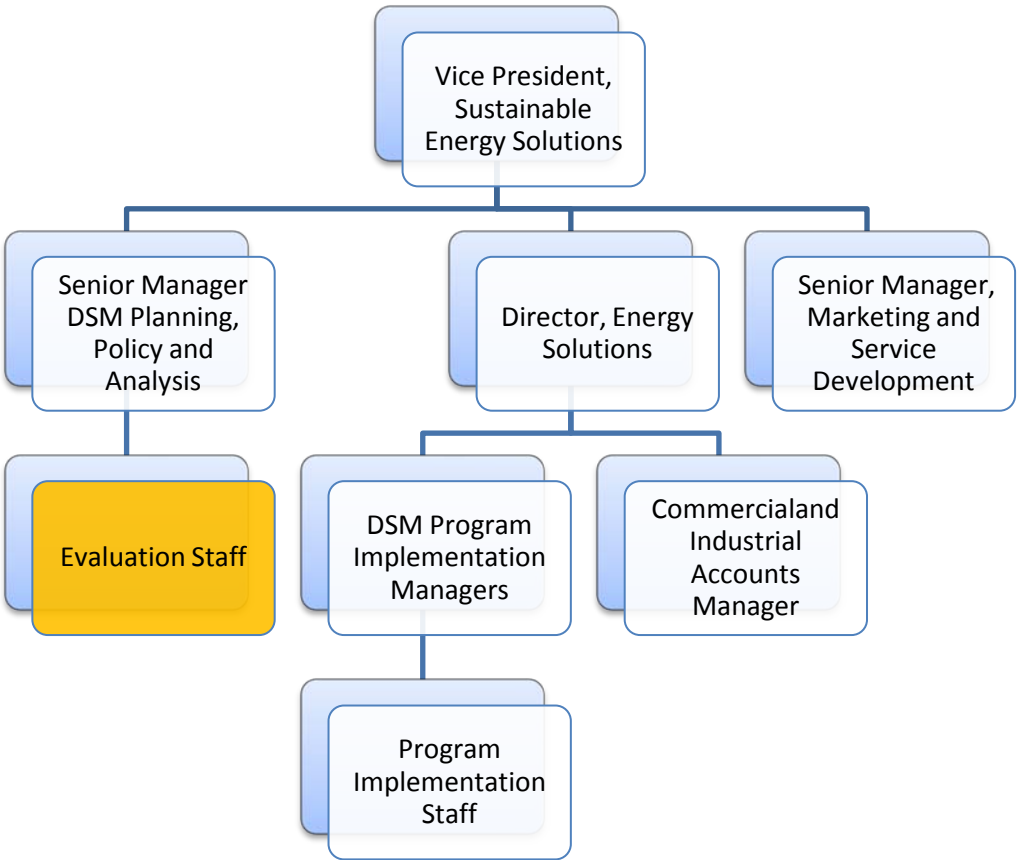
Figure A-1: PSE Simplified Org Chart



Avista

Avista is a Washington-based investor owned utility. A recent reorganization separated Avista's evaluation function from the program implementation group. While the evaluation function remained under the Vice President for Sustainable Energy Solutions, evaluation staff now report to a different senior manager than program implementation staff. Four staff members will work under the Senior Manager for DSM Planning, Policy and Analysis. While the senior manager and all four staff members carry out evaluation activities, they also have a variety of other responsibilities. The senior manager in charge of evaluation and evaluation staff works closely with program implementation staff to carry out other functions important to the organization, including annual reporting on energy efficiency, cost effectiveness analyses, the annual business planning process, high level budgeting, preparing regulatory filings, and managing stakeholder involvement in DSM activities.

Figure A-2: Avista Simplified Org Chart

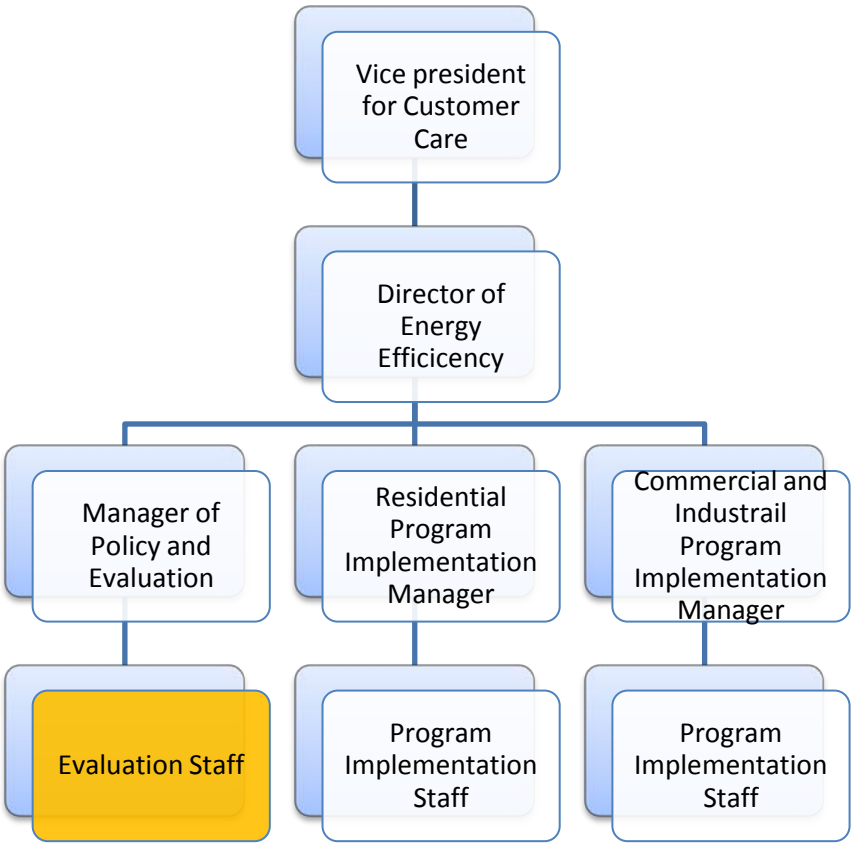


NSTAR

NSTAR is an investor-owned utility in Massachusetts. At NSTAR, a Manager of Policy and Evaluation oversees the utility's evaluation activities, as well as planning, policy, and support functions related to energy efficiency. A total of ten staff members report to the Manager of Policy and Evaluation, five of whom are directly involved in evaluation. Evaluation staff members also manage NSTAR's market research related to efficiency. The Manager of Policy and Evaluation reports to NSTAR's Director of Energy Efficiency, as do the two managers who oversee NSTAR's efficiency program implementation activities. All of NSTAR's energy efficiency activities take place under the Vice President for Customer Care.

According to NSTAR staff, the structure of the organization's efficiency functions reflects a natural division of energy efficiency activities. One manager is in charge of residential program implementation; one manager is in charge of commercial and industrial program implementation; and the Manager of Policy and Evaluation oversees planning, evaluation, and other support functions. The combination of planning, policy, regulatory affairs, evaluation, and support under a single manager reflects the relationship between these areas.

Figure A-3: NSTAR Simplified Org Chart

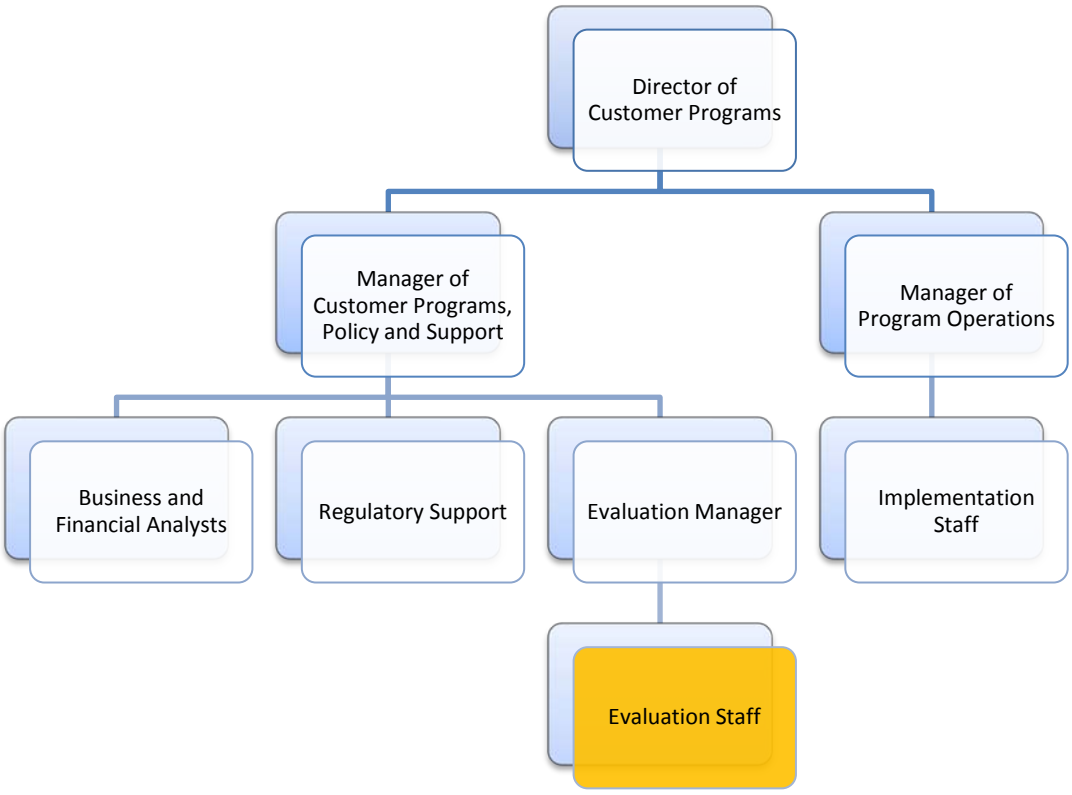


San Diego Gas & Electric

San Diego Gas & Electric is a California-based investor owned utility. SDG&E's energy efficiency activities are housed in the utility's Customer Programs area, overseen by the Director of Customer Programs. Evaluation staff report to the Manager of Customer Programs, Policy and Support, who also manages business and financial analysts and other duties related to the regulatory requirements surrounding energy efficiency and demand response. The evaluation manager oversees seven staff members, although their activities are divided between evaluation and duties related to cost effectiveness and reporting. Program implementation occurs within the Customer Programs area, although program implementation staff report to a separate manager at the same organizational level as the Manager of Customer Programs, Policy and Support.

Although SDG&E's resource planning function relies on energy efficiency program evaluation staff for information about achieved and projected energy savings, resource planning takes place outside of SDG&E's Customer Solutions area. Additionally, with the majority of its resources devoted to meeting efficiency and demand response goals established by California regulators, SDG&E conducts relatively little market research related to its efficiency programs. Evaluation staff may assist with market research that does occur, but it is not one of their primary duties, nor does it fall under the areas that the Manager of Customer Programs, Policy and Support oversees.

Figure A-4: SDG&E Simplified Org Chart



SDG&E's evaluation group has had its current structure for approximately four years. Previously, evaluation staff did not have an individual supervisor; staff members reported directly to the Manager of Customer Programs, Policy and Support. In addition to gaining a dedicated supervisor, evaluation staff took on reporting duties, which they had not previously had, and began evaluating demand response programs, which were new to SDG&E.

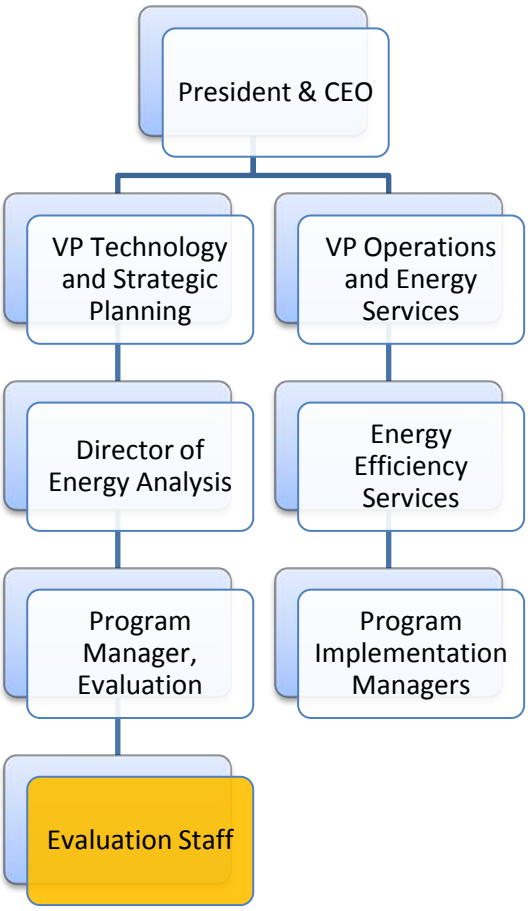
The evaluation group's place within SDG&E's energy efficiency structure reflects an effort to maintain separation between the evaluation group and program implementation staff. SDG&E management considered embedding evaluation staff under program managers, but determined that maintaining a separate, specialized evaluation group is more efficient and allows evaluators to more effectively consider findings at a portfolio level. In addition, SDG&E evaluation staff feel that organizational separation between evaluation staff and program implementation staff allows evaluation staff greater freedom to present evaluation findings that may be critical of a program or process. According to one staff member, "If you are going to do M&V right, you have to be able to tell people that the program is not working, or it's not working right."

New York State Energy Research and Development Authority (NYSERDA)

NYSERDA is the New York State authority responsible for managing the funds collected through a statewide system benefits charge. NYSERDA is also the state's energy office and, as such, is responsible for a variety of other initiatives, including research and development. NYSERDA's evaluation staff is housed in the organization's Energy Analysis program. In addition to program evaluation responsibilities, Energy Analysis performs a variety of planning, modeling and forecasting activities and provides policy analysis and regulatory support to NYSERDA as a whole. As a sub-group of Energy Analysis, the evaluation staff are expected to work closely with their peers to ensure that analytical information is incorporated into program design considerations and that evaluation activities meet the information needs of program staff and the organization as a whole. The Energy Analysis group reports to the Vice President for Technology and Strategic Planning, while program implementation staff report to the Vice President for Operations and Energy Services.

Market transformation has long been a focus of NYSERDA programs; as a result, the organization includes market characterization and assessment as an evaluation focus alongside process and impact evaluation. Evaluation staff oversee market characterization research, but NYSERDA's marketing department conducts market research to support the program-specific marketing and outreach plans being developed for new programs under New York State's Energy Efficiency Portfolio Standard. This research involves large scale benchmarking studies to identify levels of awareness and involvement with NYSERDA among both program participants and non-participants. These studies seek to identify the most effective messages and outreach strategies to increase awareness and motivate participation. Evaluation staff work closely with marketing staff in these research efforts.

Figure A-5: NYSERDA Simplified Org Chart



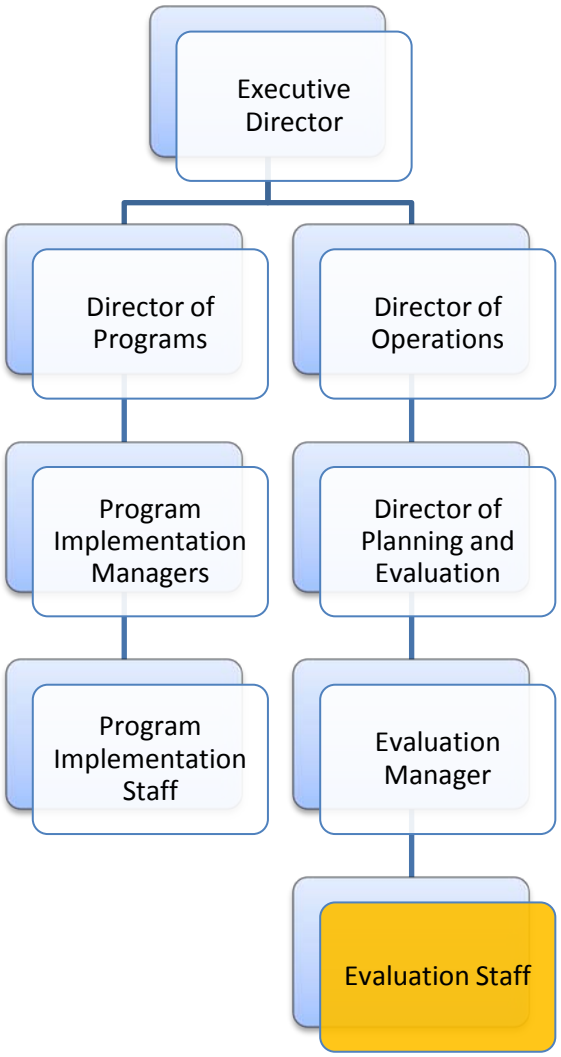
Energy Trust

Energy Trust of Oregon is the non-governmental organization responsible for managing the funds collected to support energy efficiency and renewable energy through Oregon's system benefits charge. Energy Trust's organizational structure is broadly divided into an operations unit, a programs unit, finance, and a group focused on contracting and legal issues. The programs unit implements energy efficiency and renewable energy programs; the operations unit supports the programs unit by providing a variety of support including planning and evaluation, communications and customer service, and information technology. The Energy Programs Director oversees efficiency program staff, while the Director of Operations oversees evaluation.

An Evaluation Manager reports to the Director of Planning and Evaluation, who oversees both evaluation and resource planning activities. In addition to program evaluation, the Evaluation Manager oversees a large portion of Energy Trust's market research work (the planning group and individual programs also conduct some market research), as well as annual surveys of staff and trade allies.

Energy Trust's planning and evaluation group became part of the operations unit in late 2008, after an organizational redesign led to two new positions: Director of Operations and Energy Programs Director. The evaluation group's location in the organization is designed to allow for substantial cooperation between evaluation and resource planning staff while maintaining evaluation staff's independence from program administration.

Figure A-6: Energy Trust of Oregon: Simplified Org Chart



Pennsylvania Public Utility Commission

The Pennsylvania PUC is the regulatory body for Pennsylvania's investor-owned utilities. The Pennsylvania PUC is newly charged with ensuring that utilities evaluate their energy efficiency

programs in a regular, consistent manner. The PUC is requiring utilities to hire third-party evaluation contractors is expected to create a degree of separation between program implementation staff and evaluators and bring expertise to program evaluation efforts that utility staff alone may not possess. However, the PUC remains aware of potential conflicts of interest in this structure - namely that with the utilities as their clients, third-party evaluators may be motivated to present findings in a way that reflects favorably on the utility. To mitigate the effect of this, the PUC established a State Wide Evaluator (SWE) role and hired an independent evaluation consultant to an expert advisor and ensure that evaluation results are valid. The services of the SWE are valued because commission staff members may not have the technical expertise to assess the validity of evaluation findings.

B EVALUATION UTILITY FACTORS

Table B-1, below, lists various characteristics of the evaluators and evaluation users, the context, and the specific requirements of an evaluation that influence the use of evaluations. These factors are presented to demonstrate the complexity and constraints that evaluators must be effectively able to negotiate to ensure an evaluation is used and useful to the various potential audiences and stakeholders.

Table Set B-1: Compilation of Different Factors that Influence Evaluation Use

FACTOR TYPE	CHARACTERISTICS INFLUENCING USE: PEOPLE INVOLVED
Evaluator Characteristics	Willingness to involve users • Dedication of time to foster understanding, trust with users • Choice of role as evaluation facilitator in which users lead process • Ability to create open, trusting, safe environment for all to share views and engage in learning Sensitivity to political, social, or cultural differences Demonstration of personal credibility with evaluation users Professional or personal background
User Characteristics	Identities represented: • Range of organizations, positions, and levels of professional experience • Prior training in evaluation • Level and type of education Interest in the evaluation: • Views of project, advocacy for program • Prior experience with evaluation, interest in evaluation processes • Familiarity with evaluation methods, familiarity with qualitative and quantitative data • Motivation for participation Commitment to use; use of data for decision making Professional characteristics • Openness to new ideas and changes • Willingness to speak openly • Risk tolerance Attention to quality and details Preferences in presentation of information

FACTOR TYPE	INFLUENCING USE: EVALUATION PROCESS
Evaluation Procedures	Appropriateness, rigor, sophistication of methods used Evaluation goals, questions specified according to mandates Use of a general model, criteria used in selection of model
Information Dialogue	Amount and quality of interaction between evaluator and users: • Direct communication of users' needs for evaluation information • Amount and level of communication and information dissemination between levels of bureaucracy Importance placed on and extent to which time is dedicated for reflection on evaluation process Awareness of evaluation goals, questions How unanticipated information is dealt with
Substance of Evaluation Information	Substance, relevance, specificity of evaluation information for users: • Policy maker: information on resource allocation, implementation, overall effectiveness • Program manager: information on implementation, program element effectiveness, and overall effectiveness
Evaluation Reporting	Frequency of information provided Timing information to be considered before program and policy decisions Use of jargon in and clarity of presentations and reports Mix of statistical and narrative data Information included and excluded; suppression of information

FACTOR TYPE	CHARACTERISTICS INFLUENCING USE: ENVIRONMENT & CONTEXT
Pre-Existing Evaluation Bounds	Regulatory and other written requirements Contractual obligations Fiscal constraints and budget prioritization
Organizational Features	Intra-organizational dynamics: • Role of management: management verbal support for evaluation; provision of incentives for participation; expectation of learning from evaluation process and results; recognition or reward for implementation of evaluation recommendations • Interrelationships among program unit and management • Group understanding of role in evaluation process • Program unit level of autonomy • Perceived institutional risk • Organizational resistance or open-mindedness • Inter- and intra-organizational rivalries or power struggles • Competing sources of information and opinions • Political orientation of commissioners of evaluation Degree of organizational stability; turn-over of evaluation users Organizational support for previous evaluation work: • Location and ownership of evaluation function • Extent to which organization's culture supports ongoing learning • Extent to which organization's culture supports developing evaluation capacity External factors: • External demands, threats, constraints • Dependence of decision makers on external sponsors • Role or influence of other agencies • Policy changes
Project or Program Characteristics	Age, maturity Innovativeness Overlap with other projects Visibility of program and evaluation

References

Alkin, M. (1985). *A Guide for Evaluation Decision Makers.* Beverly Hills, CA: Sage.

Alkin, M. & Taut, S. (2003). Unbundling Evaluation Use. *Studies in Educational Evaluation,* 29, 1-12.

Cousins, J.B. & Leithwood, K.A. (1986). Current Empirical Research on Evaluation Utilization. *Review of Educational Research,* 56(3), 331-364.

Leviton, L. & Hughes, E. (1981). Research on Utilization of Evaluation: A Review and Synthesis. *Evaluation Review,* 5(4), 525-548.

Patton, M. Q. (1997). *A New Century text (3rd edition).* Thousand Oaks: Sage.

Preskill, H., Zuckerman, B., Matthews, B. (2003). An Exploratory Study of the Variables Affecting Process Use. *American Journal of Evaluation,* 24(4), 423-442.

Shula, L. & Cousins, J. (1997). Evaluation Use: Theory, Research and Practice Since 1986, Evaluation Practice, 18(3), p. 195-208.

C EVALUATION GLOSSARY

Glossary

Below are some useful definitions applicable to energy program evaluation activities. These definitions were taken from the *California Energy Efficiency Evaluation Protocols.*[18]

COEFFICIENT OF VARIATION - The sample standard deviation divided by the sample mean (cv = sd/y). See page 320 of the *Evaluation Framework*.

DEMAND SAVINGS - The reduction in the demand from the pre-retrofit baseline to the post-retrofit demand, once independent variables (such as weather or occupancy) have been adjusted for. This term is usually applied to billing demand, to calculate cost savings or to peak demand, for equipment sizing purposes.

ENERGY SAVINGS - The reduction in use of energy from the pre-retrofit baseline to the post-retrofit energy use, once independent variables (such as weather or occupancy) have been adjusted for.

EVALUATION - The performance of studies and activities aimed at determining the effects of a program; any of a wide range of assessment activities associated with understanding or documenting program performance or potential performance, assessing program or program-related markets and market operations; any of a wide range of evaluative efforts including assessing program-induced changes in energy efficiency markets, levels of demand or energy savings and program cost-effectiveness.

EX-ANTE SAVINGS ESTIMATE – Administrator-forecasted savings used for program and portfolio planning purposes as filed with regulators, from the Latin for "beforehand."

EX-POST EVALUATION ESTIMATED SAVINGS – Used in California to differentiate between evaluation as versus program reported ex-post savings; Ex-post Evaluation Estimated Savings are reported by the independent evaluator after the energy impact evaluation and the associated M&V efforts have been completed. If only the term "ex-post savings" is used, California regulators assume the term is referring to this ex-post evaluation estimate, the most common usage, from the Latin for "from something done afterward."

[18] TecMarket Works Team. (2006) California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals. California Public Utilities Commission, San Francisco, CA.

EX-POST (PROGRAM) ADMINISTRATOR-ESTIMATED SAVINGS - Savings estimates reported by the Administrator after program implementation has begun (Administrator-reported ex post), from the Latin for "from something done afterward."

EX-POST (PROGRAM) ADMINISTRATOR-FORECASTED SAVINGS – Savings estimates forecasted by the Administrator during the program and portfolio planning process, from the Latin for "from something done afterward."

GROSS LOAD IMPACT - The change in energy consumption and/or demand that results directly from program-related actions taken by participants in a DSM program, regardless of why they participated. Related to Gross Energy Impact and Gross Demand Protocols.

IMPACT EVALUATION - Used to measure the program-specific induced changes in energy and/or demand usage (such kWh, kW and therms) and/or behavior attributed to energy efficiency and demand response programs.

INDIRECT ENERGY SAVINGS (INDIRECT PROGRAM ENERGY SAVINGS) - The use of the words "indirect savings" or "indirect program savings" refers to programs that are typically information, education, marketing or outreach programs in which the program's actions are expected to result in energy savings achieved through the actions of the customers exposed to the program's efforts, without direct enrollment in a program that has energy savings goals.

INTERNATIONAL PERFORMANCE MEASUREMENT AND VERIFICATION PROTOCOL (IPMVP) – The IPMVP provides an overview of current best practice techniques available for verifying results of energy efficiency, water efficiency, and renewable energy projects in commercial and industrial facilities. It may also be used by facility operators to assess and improve facility performance. The IPMVP is the leading international standard in M&V protocols. It has been translated into 10 languages and is used in more than 40 countries.

LOAD IMPACT - Changes in electric energy use, electric peak demand or natural gas use.

LOGIC MODEL - The graphical representation of the program theory showing the flow between activities, their outputs, and subsequent short-term, intermediate, and long-term outcomes. Often the logic model is displayed with these elements in boxes and the causal flow being shown by arrows from one to the others in the program logic. It can also be displayed as a table with the linear relationship presented by the rows in the table.

MARKET ASSESSMENT - An analysis function that provides an assessment of how and how well a specific market or market segment is functioning with respect to the definition of well-functioning markets or with respect to other specific policy objectives. Generally includes a characterization or description of the specific market or market segments, including a description of the types and number of buyers and sellers in the market, the key actors that influence the market, the type and number of transactions that occur on an annual basis and the extent to which energy efficiency is considered an important part of these transactions by market participants. This analysis may also include an assessment of whether or not a market has been sufficiently

transformed to justify a reduction or elimination of specific program interventions. Market assessment can be blended with strategic planning analysis to produce recommended program designs or budgets. One particular kind of market assessment effort is a baseline study, or the characterization of a market before the commencement of a specific intervention in the market, for the purpose of guiding the intervention and/or assessing its effectiveness later.

MARKET EFFECT - A change in the structure or functioning of a market or the behavior of participants in a market that result from one or more program efforts. Typically these efforts are designed to increase in the adoption of energy-efficient products, services or practices and are causally related to market interventions.

MEASURED SAVINGS - Savings or reductions in billing determinants, which are determined using engineering analysis in combination with measured data or through billing analysis.

MEASUREMENT AND VERIFICATION (M&V) - Impact evaluation will often employ metering, monitoring and verification tools to help accurately estimate the ex-post program savings. These efforts are typically referred to as "M&V," meaning either: Measurement and Verification or Monitoring and Verification, depending on the publications or reference used. M&V approaches typically are some form of field measurements taken to help identify how much energy is used before the program actions are taken, how much energy is being used after the actions are taken, the use conditions associated with an installed technology, or a change in behaviors that is to produce the energy savings.

NET LOAD IMPACT - The total change in load that is attributable to the utility DSM program. This change in load may include, implicitly or explicitly, the effects of free-drivers, free-riders, state or federal energy efficiency standards, changes in the level of energy service and natural change effects.

POWER ANALYSIS - A power analysis, executed when a study is being planned, is used to anticipate the likelihood that the study will yield a significant effect and is based on the same factors as the significance test itself. Specifically, the larger the effect size used in the power analysis, the larger the sample size; the larger (more liberal) the criterion required for significance (alpha), the higher the expectation that the study will yield a statistically significant effect. The probability-value (p-value) provided by the significance test and used to reject the null hypothesis, is a function of three factors: size of the observed effect (e.g., gross energy savings), sample size and the criterion required for significance (alpha, the level of confidence).

These three factors, together with power, form a closed system – once any three are established, the fourth is completely determined. The goal of power analysis is to find an appropriate balance among these factors by taking into account the substantive goals of the study and the resources available to the researcher.

PROCESS EVALUATION - A systematic assessment of an energy efficiency program for the purposes of documenting program operations at the time of the examination, and identifying and recommend improvements that can be made to the program to increase the program's efficiency

or effectiveness for acquiring energy resources while maintaining high levels of participant satisfaction.

SAVINGS MEASUREMENT APPROACH - The estimation of energy and demand savings associated with an energy efficiency measure for a piece of equipment, a subsystem or a system. The estimated savings are based on some kind of measured data from before and after the retrofit and may be calculated using a variety of engineering techniques.

SIMPLIFIED ENGINEERING MODEL - Engineering equations used to calculate energy usage and/or savings. These models are usually based on a quantitative description of physical processes that describe the transformation of delivered energy into useful work such as heat, lighting or motor drive. In practice, these models may be reduced to simple equations that calculate energy usage or savings as a function of measurable attributes of customers, facilities or equipment (e.g., lighting use = watts X hours of use). These models do not incorporate billing data and do not produce estimates of energy savings to which tests of statistical validity can be applied.