Table 7-6. Generic Evaluation Considerations

Overall Goals	
Does the evaluation address the key policy, regulatory, and oversight needs for evaluation information?	
Will the program success in meeting energy, demand, and emissions goals be quantifiably evaluated in the same manner as they are defined for the program?	
Does the evaluation plan represent a reasonable approach to addressing the information needs?	
Are there missing opportunities associated with the evaluation approach that should be added or considered? Are any additional co-benefits being evaluated?	
Does the impact evaluation provide the data needed to inform other evaluations that may be performed, particularly cost-effectiveness analyses?	
Has a balance been reached between evaluation costs, uncertainty of results, and value of evaluation results?	
Uncertainty of Evaluation Results	
Can the confidence and precision of the evaluation results be quantified? If so, how?	
Are there key threats to the validity of the conclusions? Are they being minimized given budget constraints and study tradeoffs? Will they be documented and analyzed?	
Is the evaluation capable of providing reliable conclusions on energy and other impacts?	
Budget, Timing, and Resources	
Does the evaluation take advantage of previous evaluations and/or concurrent ones for other programs?	
Does the cost of the study match the methods and approaches planned?	
Do the scheduled start and end times of the evaluation match the need for adequate data gathering, analysis, and reporting?	
Are adequate human resources identified?	
Does the evaluation rely on data and project access that are reasonably available?	
Reporting	
Are the time frames and scopes of evaluation reported defined?	
Do the data collection, analysis, and quality control match the reporting needs?	
Are the persistence of savings and avoided emissions being evaluated?	
Have measurement and impacts (emissions) boundaries been properly set?	
Sampling and Accuracy	
Is the sampling plan representative of the population served?	
Is the sampling plan able to support the evaluation policy objectives?	
Are there threats to the validity of the evaluation results that are incorporated into the evaluation design?	

Ameren Ex 7.2

7.4 Notes

1. A companion National Action Plan document that addresses program planning is the *Guide to Resource Planning with Energy Efficiency*, available at www.epa.gov/eeactionplan.

2. In early replacement projects, a consideration in whether to use existing conditions or code requirements for a baseline is if the replaced equipment or systems had a remaining lifetime shorter that the time period of the evaluation. In this situation, the first year(s) of the evaluation might have an existing condition baseline and the later years a code requirements baseline.

3. The CMVP program is a joint activity of the Efficiency Valuation Organization and the Association of Energy Engineers (AEE). It is accessible through EVO's Web site, <http://www.evo-world.org>.

Appendix A: National Action Plan for Energy Efficiency Leadership Group

Co-Chairs

Marsha Smith
Commissioner, Idaho Public Utilities Commission
President-Elect, National Association of Regulatory Utility Commissioners

James E. Rogers
Chairman, President, and C.E.O.
Duke Energy

Leadership Group

Barry Abramson
Senior Vice President
Servidyne Systems, LLC

Tracy Babbidge
Director, Air Planning
Connecticut Department of Environmental Protection

Angela S. Beehler
Director of Energy Regulation
Wal-Mart Stores, Inc.

Jeff Bladen
General Manager, Market Strategy
PJM Interconnection

Sheila Boeckman
Manager of Business Operations and Development
Waverly Light and Power

Bruce Braine
Vice President, Strategic Policy Analysis
American Electric Power

Cheryl Buley
Commissioner
New York State Public Service Commission

Jeff Burks
Director of Environmental Sustainability
PNM Resources

Kateri Callahan
President
Alliance to Save Energy

Jorge Carrasco
Superintendent
Seattle City Light

Lonnie Carter
President and C.E.O.
Santee Cooper

Gary Connett
Manager of Resource Planning and Member Services
Great River Energy

Larry Downes
Chairman and C.E.O.
New Jersey Natural Gas (New Jersey Resources Corporation)

Roger Duncan
Deputy General Manager, Distributed Energy Services
Austin Energy

Angelo Esposito
Senior Vice President, Energy Services and Technology
New York Power Authority

Jeanne Fox
President
New Jersey Board of Public Utilities

Anne George
Commissioner
Connecticut Department of Public Utility Control

Dian Grueneich
Commissioner
California Public Utilities Commission

Blair Hamilton
Policy Director
Vermont Energy Investment Corporation

Leonard Haynes
Executive Vice President, Supply Technologies, Renewables, and Demand Side Planning
Southern Company

Mary Healey
Consumer Counsel for the State of Connecticut
Connecticut Consumer Counsel

Joe Hoagland
Vice President, Energy Efficiency and Demand Response
Tennessee Valley Authority

Sandy Hochstetter
Vice President, Strategic Affairs
Arkansas Electric Cooperative Corporation

Helen Howes
Vice President, Environment, Health and Safety
Exelon

Bruce Johnson
Director, Energy Management
Keyspan

Mary Kenkel
Consultant, Alliance One
Duke Energy

Ruth Kiselewich
Director, Conservation Programs Unit
Baltimore Gas and Electric

Rick Leuthauser
Manager of Energy Efficiency
MidAmerican Energy Company

Harris McDowell
Senator
Delaware General Assembly

Mark McGahey
Manager
Tristate Generation and Transmission Association, Inc.

Ed Melendreras
Vice President, Sales and Marketing
Entergy Corporation

Janine Migden-Ostrander
Consumers' Counsel
Office of the Ohio Consumers' Counsel

Michael Moehn
Vice President, Corporate Planning
Ameren Services

Fred Moore
Director Manufacturing & Technology, Energy
The Dow Chemical Company

Richard Morgan
Commissioner
District of Columbia Public Service Commission

Brock Nicholson
Deputy Director
Division of Air Quality
North Carolina Air Office

Pat Oshie
Commissioner
Washington Utilities and Transportation Commission

Douglas Petitt
Vice President,
Government Affairs
Vectren Corporation

Bill Prindle
Deputy Director
American Council for an Energy-Efficient Economy

Phyllis Reha
Commissioner
Minnesota Public Utilities Commission

Roland Risser
Director, Customer Energy Efficiency
Pacific Gas and Electric

Gene Rodrigues
Director, Energy Efficiency
Southern California Edison

Art Rosenfeld
Commissioner
California Energy Commission

Gina Rye
Energy Manager
Food Lion

Jan Schori
General Manager
Sacramento Municipal Utility District

Ted Schultz
Vice President,
Energy Efficiency
Duke Energy

Larry Shirley
Division Director
North Carolina Energy Office

Tim Stout
Vice President, Energy Efficiency
National Grid

Deb Sundin
Director, Business Product Marketing
Xcel Energy

Paul Suskie
Chairman
Arkansas Public Service Commission

Dub Taylor
Director
Texas State Energy Conservation Office

Paul von Paumgartten
Director, Energy and Environmental Affairs
Johnson Controls

Brenna Walraven
Executive Director, National Property Management
USAA Realty Company

Devra Wang
Director, California Energy Program
Natural Resources Defense Council

J. Mack Wathen
Vice President, Regulatory Affairs
Pepco Holdings, Inc.

Mike Weedall
Vice President, Energy Efficiency
Bonneville Power Administration

Zac Yanez
Program Manager
Puget Sound

Henry Yoshimura
Manager, Demand Response
ISO New England Inc.

Dan Zaweski
Assistant Vice President of Energy Efficiency and Distributed Generation
Long Island Power Authority

Observers

Keith Bissell
Attorney
Gas Technology Institute

Rex Boynton
President
North American Technician Excellence

James W. (Jay) Brew
Counsel
Steel Manufacturers Association

Roger Cooper
Executive Vice President,
Policy and Planning
American Gas Association

Dan Delurey
Executive Director
Demand Response Coordinating Committee

Reid Detchon
Executive Director
Energy Future Coalition

Roger Fragua
Deputy Director
Council of Energy Resource Tribes

Jeff Genzer
General Counsel
National Association of State Energy Officials

Donald Gilligan
President
National Association of Energy Service Companies

Chuck Gray
Executive Director
National Association of Regulatory Utility Commissioners

Steve Hauser
President
GridWise Alliance

William Hederman
Member, IEEE-USA Energy Policy Committee
Institute of Electrical and Electronics Engineers

Marc Hoffman
Executive Director
Consortium for Energy Efficiency

John Holt
Senior Manager of Generation and Fuel
National Rural Electric Cooperative Association

Eric Hsieh
Manager of Government Relations
National Electrical Manufacturers Association

Lisa Jacobson
Executive Director
Business Council for Sustainable Energy

Kate Marks
Energy Program Manager
National Conference of State Legislatures

Joseph Mattingly
Vice President, Secretary and General Counsel
Gas Appliance Manufacturers Association

Kenneth Mentzer
President and C.E.O.
North American Insulation Manufacturers Association

Diane Munns
Executive Director, Retail Energy
Edison Electric Institute

Michelle New
Director, Grants and Research
National Association of State Energy Officials

Ellen Petrill
Director, Public/Private Partnerships
Electric Power Research Institute

Alan Richardson
President and C.E.O.
American Public Power Association

Andrew Spahn
Executive Director
National Council on Electricity Policy

Rick Tempchin
Director, Retail Distribution Policy
Edison Electric Institute

Mark Wolfe
Executive Director
Energy Programs Consortium

Facilitators

U.S. Department of Energy

U.S. Environmental Protection Agency

For More Information

Visit www.epa.gov/eeactionplan or contact:

Stacy Angel
Climate Protection Partnerships Division
Office of Air and Radiation
U.S. Environmental Protection Agency
angel.stacy@epa.gov
(202) 343-9606

Larry Mansueti
Director, State and Regional Assistance
Office of Electricity Deliverability and Energy Reliability
U.S. Department of Energy
lawrence.mansueti@hq.doe.gov
(202) 586-2588

Ameren Ex 7.2

Appendix
B: Glossary

This glossary is based primarily on three evaluation-related reference documents:

1. 2007 IPMVP
2. 2004 California Evaluation Framework
3. 2006 DOE EERE Guide for Managing General Program Evaluation Studies

In some cases, the definitions presented here differ slightly from the reference documents. This is due to discrepancies across documents and author interpretations.

Additionality: A criterion that says avoided emissions should only be recognized for project activities or programs that would not have "happened anyway." While there is general agreement that additionality is important, its meaning and application remain open to interpretation.

Adjustments: For M&V analyses, factors that modify baseline energy or demand values to account for independent variable values (conditions) in the reporting period.

Allowances: Allowances represent the amount of a pollutant that a source is permitted to emit during a specified time in the future under a cap and trade program.. Allowances are often confused with credits earned in the context of project-based or offset programs, in which sources trade with other facilities to attain compliance with a conventional regulatory requirement. Cap and trade program basics are discussed at the following EPA Web site: http://www.epa.gov/airmarkets/cap-trade/index.html.

Analysis of covariance (ANCOVA) model. A type of regression model also referred to as a "fixed effects" model.

Assessment boundary: The boundary within which all the primary effects and significant secondary effects associated with a project are evaluated.

Baseline: Conditions, including energy consumption and related emissions, that would have occurred without implementation of the subject project or program. Baseline conditions are sometimes referred to as "business-as-usual" conditions. Baselines are defined as either project-specific baselines or performance standard baselines.

Baseline period: The period of time selected as representative of facility operations before the energy efficiency activity takes place.

Bias: The extent to which a measurement or a sampling or analytic method systematically underestimates or overestimates a value.

California Measurement Advisory Council (CALMAC): An informal committee made up of representatives of the California utilities, state agencies, and other interested parties. CALMAC provides a forum for the development, implementation, presentation, discussion, and review of regional and statewide market assessment and evaluation studies for California energy efficiency programs conducted by member organizations.

Co-benefits: The impacts of an energy efficiency program other than energy and demand savings.

Coincident demand: The metered demand of a device, circuit, or building that occurs at the same time as the peak demand of a utility's system load or at the same time as some other peak of interest, such as building or facility peak demand. This should be expressed so as to indicate the peak of interest (e.g., "demand coincident with the utility system peak") Diversity factor is defined as the ratio of the sum of the demands of a group of users to their coincident maximum demand. Therefore, diversity factors are always equal to one or greater.

Comparison group: A group of consumers who did not participate in the evaluated program during the program year and who share as many characteristics as possible with the participant group.

Conditional Savings Analysis (CSA): A type of analysis in which change in consumption modeled using regression analysis against presence or absence of energy efficiency measures.

Confidence: An indication of how close a value is to the true value of the quantity in question. Confidence is the likelihood that the evaluation has captured the true impacts of the program within a certain range of values (i.e., precision).

Cost-effectiveness: An indicator of the relative performance or economic attractiveness of any energy efficiency investment or practice. In the energy efficiency field, the present value of the estimated benefits produced by an energy efficiency program is compared to the estimated total costs to determine if the proposed investment or measure is desirable from a variety of perspectives (e.g., whether the estimated benefits exceed the estimated costs from a societal perspective).

Database for Energy-Efficient Resources (DEER): A California database designed to provide well-documented estimates of energy and peak demand savings values, measure costs, and effective useful life.

Deemed savings: An estimate of an energy savings or energy-demand savings outcome (gross savings) for a single unit of an installed energy efficiency measure that (a) has been developed from data sources and analytical methods that are widely considered acceptable for the measure and purpose and (b) is applicable to the situation being evaluated.

Demand: The time rate of energy flow. Demand usually refers to electric power measured in kW (equals kWh/h) but can also refer to natural gas, usually as Btu/hr, kBtu/hr, therms/day, etc.

Direct emissions: Direct emissions are changes in emissions at the site (controlled by the project sponsor or owner) where the project takes place. Direct emissions are the source of avoided emissions for thermal energy efficiency measures (e.g., avoided emissions from burning natural gas in a water heater).

Effective useful life: An estimate of the median number of years that the efficiency measures installed under a program are still in place and operable.

Energy efficiency: The use of less energy to provide the same or an improved level of service to the energy consumer in an economically efficient way; or using less energy to perform the same function. "Energy conservation" is a term that has also been used, but it has the connotation of doing without a service in order to save energy rather than using less energy to perform the same function.

Energy efficiency measure: Installation of equipment, subsystems or systems, or modification of equipment, subsystems, systems, or operations on the customer side of the meter, for the purpose of reducing energy and/or demand (and, hence, energy and/or demand costs) at a comparable level of service.

Engineering model: Engineering equations used to calculate energy usage and savings. These models are usually based on a quantitative description of physical processes that transform delivered energy into useful work such as heat, lighting, or motor drive. In practice, these models may be reduced to simple equations in spreadsheets that calculate energy usage or savings as a function of measurable attributes of customers, facilities, or equipment (e.g., lighting use = watts × hours of use).

Error: Deviation of measurements from the true value.

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, 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: Forecasted savings used for program and portfolio planning purposes. (From the Latin for "beforehand.")

Ex post evaluation estimated savings: Savings estimates reported by an evaluator after the energy impact evaluation has been completed. (From the Latin for "from something done afterward.")

Free driver: A non-participant who has adopted a particular efficiency measure or practice as a result of the evaluated program.

Free rider: A program participant who would have implemented the program measure or practice in the absence of the program. Free riders can be total, partial, or deferred.

Gross savings: The change in energy consumption and/or demand that results directly from program-related actions taken by participants in an efficiency program, regardless of why they participated.

Impact evaluation: An evaluation of the program-specific, directly induced changes (e.g., energy and/or demand usage) attributable to an energy efficiency program.

Independent variables: The factors that affect energy use and demand, but cannot be controlled (e.g., weather or occupancy).

Indirect emissions: Changes in emissions that occur at the emissions source (e.g., the power plant). Indirect emissions are the source of avoided emissions for electric energy efficiency measures.

Interactive factors: Applicable to IPMVP Options A and B; changes in energy use or demand occurring beyond the measurement boundary of the M&V analysis.

Leakage: In the context of avoided emissions, emissions changes resulting from a project or program not captured by the primary effect (typically the small, unintended emissions consequences). Sometimes used interchangeably with "secondary effects," although leakage is a more "global" issue whereas secondary, interactive effects tend to be considered within the facility where a project takes place.

Load shapes: Representations such as graphs, tables, and databases that describe energy consumption rates as a function of another variable such as time or outdoor air temperature.

Market effect evaluation: An evaluation of the change in the structure or functioning of a market, or the behavior of participants in a market, that results from one or more program efforts. Typically the resultant market or behavior change leads to an increase in the adoption of energy-efficient products, services, or practices.

Market transformation: A reduction in market barriers resulting from a market intervention, as evidenced by a set of market effects, that lasts after the intervention has been withdrawn, reduced, or changed.

Measurement: A procedure for assigning a number to an observed object or event.

Measurement and verification (M&V): Data collection, monitoring, and analysis associated with the calculation of gross energy and demand savings from individual sites or projects. M&V can be a subset of program impact evaluation.

Measurement boundary: The boundary of the analysis for determining direct energy and/or demand savings.

Metering: The collection of energy consumption data over time through the use of meters. These meters may collect information with respect to an end-use, a circuit, a piece of equipment, or a whole building (or facility). Short-term metering generally refers to data collection for no more than a few weeks. End-use metering refers specifically to separate data collection for one or more end-uses in a facility, such as lighting, air conditioning or refrigeration. Spot metering is an instantaneous measurement (rather than over time) to determine an energy consumption rate.

Monitoring: Gathering of relevant measurement data, including but not limited to energy consumption data, over time to evaluate equipment or system performance, e.g., chiller electric demand, inlet evaporator temperature and flow, outlet evaporator temperature, condenser inlet temperature, and ambient dry-bulb temperature and relative humidity or wet-bulb temperature, for use in developing a chiller performance map (e.g., kW/ton vs. cooling load and vs. condenser inlet temperature).

Net savings: The total change in load that is attributable to an energy efficiency program. This change in load may include, implicitly or explicitly, the effects of free drivers, free riders, energy efficiency standards, changes in the level of energy service, and other causes of changes in energy consumption or demand.

Net-to-gross ratio (NTGR): A factor representing net program savings divided by gross program savings that is applied to gross program impacts to convert them into net program load impacts.

Non-participant: Any consumer who was eligible but did not participate in the subject efficiency program, in a given program year. Each evaluation plan should provide a definition of a non-participant as it applies to a specific evaluation.

Normalized annual consumption (NAC) analysis: A regression-based method that analyzes monthly energy consumption data.

Participant: A consumer that received a service offered through the subject efficiency program, in a given program year. The term "service" is used in this definition to suggest that the service can be a wide variety of services, including financial rebates, technical assistance, product installations, training, energy efficiency information or other services, items, or conditions. Each evaluation plan should define "participant" as it applies to the specific evaluation.

Peak demand: The maximum level of metered demand during a specified period, such as a billing month or a peak demand period.

Persistence study: A study to assess changes in program impacts over time (including retention and degradation).

Portfolio: Either (a) a collection of similar programs addressing the same market (e.g., a portfolio of residential programs), technology (e.g., motor efficiency programs), or mechanisms (e.g., loan programs) or (b) the set of all programs conducted by one organization, such as a utility (and which could include programs that cover multiple markets, technologies, etc.).

Potential studies: Studies conducted to assess market baselines and savings potentials for different technologies and customer markets. Potential is typically defined in terms of technical potential, market potential, and economic potential.

Precision: The indication of the closeness of agreement among repeated measurements of the same physical quantity.

Primary effects: Effects that the project or program are intended to achieve. For efficiency programs, this is primarily a reduction in energy use per unit of output.

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 recommending improvements to increase the program's efficiency or effectiveness for acquiring energy resources while maintaining high levels of participant satisfaction.

Program: A group of projects, with similar characteristics and installed in similar applications. Examples could include a utility program to install energy-efficient lighting in commercial buildings, a developer's program to build a subdivision of homes that have photovoltaic systems, or a state residential energy efficiency code program.

Project: An activity or course of action involving one or multiple energy efficiency measures, at a single facility or site.

Rebound effect: A change in energy-using behavior that yields an increased level of service and occurs as a result of taking an energy efficiency action.

Regression analysis: Analysis of the relationship between a dependent variable (response variable) to specified independent variables (explanatory variables). The mathematical model of their relationship is the regression equation.

Reliability: Refers to the likelihood that the observations can be replicated.

Reporting period: The time following implementation of an energy efficiency activity during which savings are to be determined.

Resource acquisition program: Programs designed to directly achieve energy and or demand savings, and possibly avoided emissions

Retrofit isolation: The savings measurement approach defined in IPMVP Options A and B, and ASHRAE Guideline 14, that determines energy or demand savings through the use of meters to isolate the energy flows for the system(s) under consideration.

Rigor: The level of expected confidence and precision. The higher the level of rigor, the more confident one is that the results of the evaluation are both accurate and precise.

Secondary effects: Unintended impacts of the project or program such as rebound effect (e.g., increasing energy use as it becomes more efficient and less costly to use), activity shifting (e.g., when generation resources move to another location), and market leakage (e.g., emission changes due to changes in supply or demand of commercial markets). These secondary effects can be positive or negative.

Spillover: Reductions in energy consumption and/or demand caused by the presence of the energy efficiency program, beyond the program-related gross savings of the participants. There can be participant and/or non-participant spillover.

Statistically adjusted engineering (SAE) models: A category of statistical analysis models that incorporate the engineering estimate of savings as a dependent variable.

Stipulated values: See "deemed savings."

Takeback effect: See "rebound effect."

Uncertainty: The range or interval of doubt surrounding a measured or calculated value within which the true value is expected to fall within some degree of confidence.

Ameren Ex 7.2

Appendix C: Other Evaluation Types

C.1 Process, Market Effects, and Cost-Effectiveness Evaluations

The following subsections briefly introduce two other, non-impact types of evaluations and cost-effectiveness analysis. These types of evaluations can involve interrelated activities and have interrelated results, and are often conducted at the same time. Table C-1 compares these three types plus impact evaluations.

C.1.1 Process Evaluations

The goal of process evaluations is to produce improved and more cost-effective programs. Thus, process evaluations examine the efficiency and effectiveness of program implementation procedures and systems. These evaluations usually consist of asking questions of those involved in the program, analyzing their answers, and comparing results to established best practices.

Process evaluations are particularly valuable when:

- The program is new or has many changes.
- Benefits are being achieved more slowly than expected.
- There is limited program participation or stakeholders are slow to begin participating.
- The program has a slow startup.
- Participants are reporting problems.
- The program appears not to be cost-effective.

Typical process evaluation results involve recommendations for changing a program's structure, implementation approaches, or program design, delivery, and goals.

The primary mechanism of process evaluations is data collection (e.g., surveys, questionnaires, and interviews) from administrators, designers, participants (such as facility operators), implementation staff (including contractors, subcontractors, and field staff), and key policy makers. Other elements of a process evaluation can include workflow and productivity measurements; reviews, assessments, and testing of records, databases, program-related materials, and tools; and possibly

Table C-1. Program Evaluation Types

Evaluation Type	Description	Uses
Impact Evaluation	Quantifies direct and indirect benefits of the program.	Determines the amount of energy and demand saved, the quantity of emissions reductions, and possibly the co-benefits.
Process Evaluations	Indicates how the program implementation procedures are performing from both administration and participant perspectives.	Identifies how program processes can be improved.
Market Effects Evaluation	Indicates how the overall supply chain and market have been affected by the program.	Determines changes that have occurred in markets and whether they are sustainable with or without the program.
Cost-Effectiveness Evaluation	Quantifies the cost of program implementation and compares with program benefits.	Determines whether the energy efficiency program is a cost-effective investment as compared to other programs and energy supply resources.

Table C-2. Elements of a Typical Process Evaluation

• Program Design – The program mission – Assessment of program logic – Use of new practices or best practices	• Program Administration – Program oversight – Program staffing – Management and staff training – Program information and reporting
• Program Implementation – Quality control – Operational practice—how program is implemented – Program targeting, marketing, and outreach efforts – Program timing	• Participant Response – Participant interaction and satisfaction – Market and government allies interaction and satisfaction

collection and analysis of relevant data from third-party sources (e.g., equipment vendors, trade allies).

Table C-2 lists examples of the issues that are typically assessed during a process evaluation.

C.1.2 Market Effects Evaluations

Program-induced changes that affect non-participants or the way a market operates are addressed in market effects evaluations. One way to think of these is that they estimate the effect a program has on future energy efficiency activities.

Market effects evaluations often involve a significant undertaking, since they are designed to determine whether the market is changing. For example, a market effects study could evaluate increases in the adoption of the products or services being promoted by the program (or more likely, a portfolio of programs). It might answer the question: Are vendors stocking and promoting more energy efficiency technologies as a result of the program? Market effects are sometimes called the ultimate test of a program's success, answering the question—will efficiency best practices continue in the marketplace, even after the current program ends?

Potential Studies

Another form of market study is called a potential study. Potential studies are conducted before a program is implemented in order to assess market baselines and savings potentials for different technologies and customer markets. These studies can also assess customer needs and barriers to adoption of energy efficiency, as well as how best to address these barriers through program design. Potential studies indicate what can be expected in terms of savings from a program. Potential is often defined in terms of *technical potential* (what is technically feasible given commercially available products and services), *economic potential* (which is the level of savings that can be achieved assuming a certain level of participant and/or societal cost-effectiveness is required), and *market potential* (what the market can provide, which is almost always less than market potential). Findings also help managers identify the program's key markets and clients and how to best serve the intended customers.

Market effects evaluations usually consist of surveys, reviews of market data, and analysis of the survey results and collected data. Some possible results from a market assessment include:

- Total market effects.
- An estimate of how much of the market effect is due to the program being evaluated.
- An estimate of whether the market effect is sustainable.

A market effects evaluation analyzes:

- Are the entities that undertook efficiency projects undertaking additional projects or incorporating additional technologies in their facilities that were not directly induced by the program? This might indicate that the facility operators have become convinced of the value of, for example, high-efficiency motors, and are installing them on their own.
- Are entities that did not undertake projects now adopting concepts and technologies that were encouraged by the program? This might indicate that the program convinced other facility operators of the advantages of the efficiency concepts.
- Are manufacturers, distributors, vendors, and others involved in the supply chain of efficiency products (and services) changing their product offerings, how they are marketing them, how they are pricing them, stocking them, etc.? The answers can indicate how the supply chain is adapting to changes in supply of and demand for efficiency products.

As can be deduced, the market effects evaluation can easily overlap with the spillover analyses conducted as part of an impact evaluation. Market effects studies, however, are interested in long-term, sustained effects, versus a more short-term spillover perspective. According to a study by the New England Efficiency Partnership (NEEP, 2006), most programs use direct participation and spillover as the basis for estimating market transformation program benefits, rather than projections of baselines and market penetration. Anecdotal evidence suggests that measurement of participant spillover is relatively common, while measurement of non-participant spillover is inconsistent across program administrators.

About one fourth of the states in the 2006 study estimated ultimate effects by projecting change in market penetration relative to a projected baseline for at least some of their market transformation programs.

C.1.3 Cost-Effectiveness Analyses

Cost-effectiveness (sometime called cost-benefit) evaluations compare program benefits and costs, showing the relationship between the value of a program's outcomes and the costs incurred to achieve those benefits. The findings help program managers judge whether to retain, revise, or eliminate program elements and provide feedback on whether efficiency is a wise investment as compared to energy generation and/or procurement options. It also often a key component of the evaluation process for programs using public or utility ratepayer funds.

A variety of frameworks have historically been used to assess cost-effectiveness of energy efficiency initiatives. In the late 1970s, CPUC implemented a least-cost planning strategy in which demand-side reductions in energy use were compared to supply additions. One result of this strategy was *The Standard Practice Manual* (SPM). This document provided several methodologies for conducting cost-benefit analyses of utility-administered efficiency programs. The first version of the SPM was published in 1983. The document has been updated from time to time, with the most reason version dated 2001 (California State Governor's Office, 2001). The SPM is perhaps the definitive resource for information on cost-effectiveness tests for efficiency programs.

The SPM established several tests that can be used to evaluate the cost-effectiveness of publicly funded energy efficiency initiatives. These include the ratepayer impact measure test, the utility cost test, the participant test, the total resource cost test, and the societal test. These metrics vary in terms of (a) their applicability to different program types, (b) the cost and benefit elements included in the calculation, (c) the methods by which the cost and benefit elements are computed, and (d) the uses of the results. Most regulated utility efficiency programs use one or more versions of these tests, sometimes with variations unique to the requirements of a particular regulatory commission. Definitions of these tests (paraphrased from the SPM) are provided below.

- **Total resource cost (TRC) test.** The TRC test measures the net costs of a demand-side management program as a resource option based on the total costs of the program, including both the participants' and the utility's costs. The TRC ratio equals the benefits of the program, in terms of value of energy and demand saved, divided by the net costs. The ratio is usually calculated on a life-cycle basis considering savings and costs that accrue over the lifetime of installed energy efficiency equipment, systems, etc. When the TRC test is used, if the ratio is greater than 1.0, then the program is considered cost-effective, with of course proper consideration of uncertainties in the TRC ratio calculation. This is probably the most commonly applied cost-effectiveness test.
- **Utility cost (UC) test.** The UC test measures the net costs of a demand-side management program as a resource option based on the costs incurred by the administrator of the program (assumed to be a utility, though it can be any organization), excluding any net costs incurred by the participant. The benefits are the same as the TRC benefits (energy and demand savings value), but the costs are defined more narrowly and do not include consumer costs.
- **Participant test.** The participant test assesses cost-effectiveness from the participating consumer's perspective by calculating the quantifiable benefits and costs to the consumer of participating in a program. Since many consumers do not base their decision to participate entirely on quantifiable variables, this test is not necessarily a complete measure of all the benefits and costs a participant perceives.
- **Societal test.** The societal test, a modified version of the TRC, adopts a societal rather than a utility service area perspective. The primary difference between the societal and TRC tests is that, to calculate life cycle costs and benefits, the societal test accounts for externalities (e.g., environmental benefits), excludes tax credit benefits, and uses a societal discount rate.
- **Ratepayer impact measure (RIM) test.** The RIM test only applies to utility programs. It measures what happens to consumer bills or rates due to changes in utility revenues and operating costs caused by the program. This test indicates the direction and magnitude of the expected change in customer bills or rate levels.

C.2 Evaluating Other Program Types

This Guide focuses on the evaluation of programs whose primary goal is to directly achieve energy and demand savings and perhaps avoided emissions—resource acquisition programs. While all efficiency programs hope to achieve savings, some are designed to achieve these savings more indirectly. Evaluation of three other common program types (market transformation, codes and standards, and education and training) is briefly discussed below.

C.2.1 Market Transformation Programs

Market transformation (MT) denotes a permanent, or at least long-term, change in the operation of the market for energy efficiency products and services. MT programs attempt to reduce market barriers through market interventions that result in documented market effects that lasts after the program (intervention) has been withdrawn reduced or changed. During the 1990s, the focus of many energy efficiency efforts shifted from resource acquisition to market transformation. Subsequently there has been a shift back; resource acquisition, MT, and other program types are now implemented, often in a complementary manner. To a large extent, all programs can be considered MT in that they involve changing how energy efficiency activities take place in the marketplace.

MT evaluation tends to be a combination of impact, process, and market effect evaluation and can also include cost-effectiveness evaluations. However, given that the ultimate aim of MT programs is to increase the adoption of energy efficient technologies and practices, MT evaluation usually focuses first on energy efficiency adoption rates by market actors and second on the directly associated energy and demand savings. Also, MT programs are dynamic, and thus the nature of market effects can be expected to vary over time. Market actors that influence end-use consumer choices include installation and repair contractors, retailer staffs, architects, design engineers, equipment distributors, manufacturers, and of course the consumers themselves.

Evaluation plays an important role in providing the kind of feedback that can be used to refine the design of market interventions. This role is equally important for resource acquisition and MT interventions, but arguably more complex for MT programs since the interest is long-term changes in the market versus more immediate and direct energy savings for resource acquisition programs. Most importantly, evaluation for MT entails the collection of information that can be used to refine the underlying program theory (see side bar).

Evaluation of MT interventions also needs to focus on the mechanism through which changes in adoptions and energy usage are ultimately induced. This means that considerable attention must be focused on indicators of market effects through market tracking. Thus, a MT evaluation might first report changes in sales patterns and volumes for particular efficiency products as an indication of program progress in meeting program goals. (For more information on MT evaluation, see DOE, 2007).

C.2.2 Codes and Standards Programs

Most codes and standards programs involve (a) new or changed building codes or appliance and equipment standards and/or (b) increasing the level of compliance with code requirements or appliance standards. These programs are intended to save energy and demand and achieve co-benefits, primarily in new construction or major retrofits (for building codes) or when new equipment is purchased (appliance and equipment standards).

The primary approach to establishing energy and demand savings (and avoided emissions) values for the codes and standards programs is to assess the energy and demand impacts of the market adoption and decision changes caused by the new, modified, or better-enforced codes or standards and then adjust those savings to account for what would have occurred if the code or standard change or enforcement did not occur. The evaluation must identify the net energy impacts that

Theory-Based Evaluation: a Guiding Principle for MT Evaluation

Theory-based evaluation (TBE), an evaluation approach that has been widely used in the evaluation of social programs in other fields, has gained some foothold in the energy efficiency industry over the past few years. It involves a relatively detailed and articulated *program theory*, established up front, that specifies the sequence of events a program is intended to cause, along with the precise causal mechanisms leading to these events. Evaluation then focuses on testing the consistency of observed events with the overall program theory.

A TBE can be considered a process of determining whether a program theory is correct or not (i.e., testing a hypothesis). For example, with an incentive program, the theory is that paying a certain level of incentives will result in a certain level of energy and demand savings.

Having well-defined program theories helps focus an evaluation objective on assessing the validity of those theories, primarily to see whether a program concept is successful and should be expanded and/or repeated.

In the energy efficiency field to date, TBE is particularly well adapted to evaluating the effectiveness of market transformation initiatives. This is largely because market transformation tends to take a relatively long time to occur, involve a relatively large number of causal steps and mechanisms, and encompass changing the behavior of multiple categories of market actors, all of which makes it particularly fruitful to focus on specifying and testing a detailed and articulated program theory.

Provided by Ralph Prahl.

can be directly attributed to the program's actions that would not have occurred over the course of the normal, non-program-influenced operations of the market. For example, analysis of a new appliance standard would involve (a) estimating the life-cycle savings associated with each new appliance placed into service as compared to a standard practice or old-standard appliances, (b) multiplying those savings by the rate over time that the new appliances are placed into service, and (c) adjusting the resulting savings estimate by the number of high-efficiency appliances that consumers would have purchased even if the standard were not in place.

C.2.3 Education and Training Programs

Education and training programs only indirectly result in energy and demand savings. They can include advertising, public service announcements, education efforts, training activities, outreach efforts, demonstration projects, and other information- or communication-based efforts. These programs may be targeted to either end-use customers or other market actors whose activities influence the energy-related choices of end-use customers.

Typically, information and education programs have one or more of the following general goals:

- Educate energy consumers regarding ways to increase the energy efficiency of their facilities and activities, and thus convince them to take actions that help them manage their consumption or adopt more energy-efficient practices.
- Inform energy consumers and/or other market actors about program participation opportunities in order to increase enrollment in these programs.
- Inform energy consumers and/or other market actors about energy issues, behaviors, or products in an effort to transform the normal operations of the market.

Almost every energy efficiency program provides some level of educational and/or informational content. However, education-specific programs are typically designed to achieve energy or demand savings indirectly through changes in behavior, over time (market transformation) or via increased enrollments in other resource acquisition programs.

Understanding and Affecting Behavior

Some recent energy efficiency program efforts have focused on understanding the behavior and decision-making of individuals and organizations with respect to the design, adoption, and use of energy efficiency actions and on using that knowledge to help accelerate the implementation of energy efficiency activities. The proceedings of the 2007 Behavior, Energy and Climate Change Conference provide information on these approaches. See <http://ciee.ucop.edu/>.

For education and training programs, evaluations focus on documenting the degree to which the programs are achieving their desired effects within the markets targeted by the program, which is educating and training people on energy efficiency. The primary mechanisms for this type of evaluation are surveys and focus groups. The following are examples of information topics that may be collected as part of surveys and focus groups (paraphrased from the *California Protocols*):

- Information and education program evaluation topics:
  - Number and percent of customers reached or made aware.
  - Number and percent of customers reached who take recommended actions.
  - Number and type of actions taken as a result of the program.
  - Changes in awareness or knowledge by topic or subject area, by type of customer targeted.
  - Customer perception of the value of the information and/or education received.
  - Elapsed time between information exposure and action(s) taken by type of customer targeted.
  - Attribution of cause for actions taken when multiple causes may be associated with the actions taken.
  - Influence of the program on dealers, contractors, and trade allies.

- Effects of the program on manufacturers and distributors.

- Training program evaluation topics:
  - Pre-program level of knowledge to compare with post-program levels.
  - The specific knowledge gained through the program.
  - The relevance and usefulness of the training as it relates to the participants' to specific needs and opportunities to use the information.
  - Future opportunities and plans for incorporating the knowledge gained into actions or behaviors that provide energy impacts.
  - Whether participants would recommend the training to a friend or colleague.
  - Participant recommendations for improving the program.

Note that programs with large training efforts, or programs designed solely for training, should have evaluation designs that are mindful of the rich literature and methods on evaluating training programs that are available from the larger evaluation community.

Ameren Ex 7.2

Appendix D: Uncertainty[1]

This appendix provides an introduction on how uncertainty is defined, as well as an overview of the range of factors that contribute to uncertainty and the impact of each of these. This discussion's target audience is evaluators who need an introduction to uncertainty and managers responsible for overseeing evaluations, such as government, regulatory agency staff, and utility staff responsible for energy efficiency evaluations. This appendix assumes readers are *not* trained statisticians and does not aim to provide the reader with all of the tools, formulas, and programs to calculate measures of uncertainty. Rather, we seek to provide the reader with a solid foundation for understanding key concepts and determining evaluation strategies for identifying and mitigating uncertainty. Finally, we wish to provide readers with the ability to review, as needed, more technical and detailed discussions of each source of uncertainty and its mitigation.

D.1 Sources of Uncertainty

Uncertainty is a measure of the "goodness" of an estimate. Without some measurement of uncertainty, it is impossible to judge an estimate's value as a basis for decision-making: uncertainty is the amount or range of doubt surrounding a measured or calculated value. Any report of gross or net program savings, for instance, has a halo of uncertainty surrounding the reported value relative to the true gross or net savings (which are not known). Defined this way, uncertainty is an overall indicator of how well a calculated or measured value represents a true value.

Program evaluation seeks to reliably determine energy and demand *savings* (and, potentially, non-energy benefits) with some reasonable accuracy. This objective can be affected by:

- **Systematic sources of error,** such as measurement error, non-coverage error, and non-response error.
- **Random error**—error occurring by chance, attributable to using a population sample rather than a census to develop the calculated or measured value.

The distinction between systematic and random error is important because different procedures are required to identify and mitigate each. The amount of random error can be estimated using statistical tools, but other means are needed for systematic error. While additional investment in the estimation process reduce both types of error, tradeoffs between evaluation costs and reductions in uncertainty are inevitably required.

D.1.1. Sources of Systematic Error

Systematic errors potentially occur from the way data are:

- **Measured.** At times, equipment used to measure consumption may not be completely accurate. Human errors (e.g., errors in recording data) can also cause this type of error. Measurement error is reduced by investing in more accurate measurement technology and more accurately recording and checking data. The magnitude of such errors is often not large enough to warrant concern in a program evaluation and is largely provided by manufacturer's specifications. In most applications, this error source is ignored, particularly when data sources are utility-grade electricity or natural gas meters. However, other types of measurements, such as flow rates in water or air distribution systems, can have significant errors.
- **Collected.** If some parts of a population are not included in the sample, non-coverage errors result, and the value calculated from the sample might not accurately represent the entire population of interest. Non-coverage error is reduced by investing in a sampling plan that addresses known coverage issues.

For instance, a survey implemented through several modes, such as phone and Internet, can sometimes address known coverage issues. Non-response errors occur when some portion or portions of the population, with different attitudes or behaviors, are less likely to provide data than are other portions. For a load research or metering study, if certain types of households are more likely to refuse to participate or if researchers are less likely to be able to obtain required data from them, the values calculated from the sample will understate the contribution of this portion of the population and over-represent the contribution of sample portions more likely to respond. In situations where the under-represented portion of the population has different consumption patterns, non-response error is introduced into the value calculated from the sample. Non-response error is addressed through investments that increase the response rate, such as incentives and multiple contact attempts.

- **Described (modeled).** Estimates are created through statistical models. Some are fairly simple and straightforward (e.g., estimating the mean), and others are fairly complicated (e.g., estimating response to temperature through regression models). Regardless, errors can occur due to the use of the wrong model, assuming inappropriate functional forms, inclusion of irrelevant information, or exclusion of relevant information. For example, in determining energy savings, a researcher may be required to adjust measured energy use data to make comparisons with a baseline. This process can introduce systematic errors.

D.1.2 Sources of Random Error

Whenever a *sample* of a population is selected to represent the population itself—whether the sample is of appliances, meters, accounts, individuals, households, premises, or organizations—there will be some amount of *random sampling error.* The sample selected is only one of a large number of possible samples of the same size and design that could have been selected from that population. For each sample, values calculated will differ from the other potential samples simply because of the element of chance in choosing particular elements. This variability is termed random sampling error. Random sampling error, unlike the systematic errors discussed above, can be estimated using statistical tools (assuming the sample was drawn randomly).

When the time savings actually take place is also essential—another layer of sampling error. Typically, what (or who) is sampled and when they are sampled (e.g., metering energy consumption over one week, metering 5 percent of impacted equipment) introduces uncertainty.

Altering sample design can reduce uncertainty from random sampling error (for instance, increasing the number of elements sampled or changing the way elements are grouped together prior to sampling). As expected, random error and sampling costs are inversely proportional in most instances.

In addition to random sampling error, random measurement error may be introduced by other factors, such as respondents' incorrectly recalling dates or expenses, or other differences in a respondent's mood or circumstances that affect how they answer a question. These other types of random measurement error are generally assumed to "even out," so that they do not affect the mean or point estimate, but only increase the variability. For this reason, researchers generally do not attempt to quantify the potential for random measurement error in the data.

D.2 Energy Efficiency Evaluation Uncertainty

The biggest challenge in evaluating energy efficiency programs is a lack of direct measurement. Energy savings are what *did not happen*, but energy consumption is actually what is measured. The difference between energy consumption and what energy consumption *would have been* had energy efficiency measures not been installed provides a measure of energy savings. Savings computation therefore involves comparing measured energy data and a calculation of "adjustments" to convert both measurements to the same set of operating conditions (i.e., a baseline). Both measurement and adjustment processes introduce uncertainty.

These processes produce statistical "estimates" with reported or expected values and some level of variability. In other words, true values cannot be known; only estimates can be made, with some level of uncertainty. Physical measurements and statistical analyses are based on estimation of central tendencies (mean, median, mode) and associated quantification of variations (standard deviation, standard error, variance).

Because uncertainty arises from many different sources, it is usually difficult to identify and quantify the effect of all potential sources. Research reports often identify only uncertainty arising from random sampling error, because this source of error is usually the easiest component to quantify. Convenient measures, such as confidence intervals and statistical significance tests, are available to provide quantitative estimates of the uncertainty. Uncertainty attributable to forms of systematic error does not have a single comparable measure to provide a parsimonious estimate of uncertainty. Rather, these sources are specific to individual studies, depending on equipment used, research staff, or research and data collection procedures employed. To assess uncertainty from systematic sources, evaluators must address the rigor of evaluation procedures.

Evaluating uncertainty is an ongoing process that can consume time and resources. It may also require the services of specialists familiar with data analysis techniques, further data collection, or additional equipment. Reducing errors usually increases evaluation costs. Thus, improved accuracy should be justified by the value of the improved information.

D.3 Statistical Terms

While studying a phenomenon at the population level (a census) produces greater accuracy, the cost is almost always prohibitive. If properly designed, samples can provide accurate estimates at a greatly reduced cost. Statistics are mathematical methods that, applied to sample data, can help make inferences about whole populations and aid decisions in the face of uncertainty.

For any value calculated from a sample, a set of descriptive statistics, such as the mean, standard deviation, standard error, and a confidence interval, can be calculated. Standard deviation is a measure of variability showing the extent of dispersion around the mean. In normally distributed data, about 68 percent of observations are within one standard deviation of the mean; so a large standard deviation indicates greater dispersion of an individual observation from each sample member, while a smaller standard deviation indicates less dispersion. Based on the amount of variability and standard deviation, a confidence interval can be calculated.

To communicate evaluation results credibly, outcomes need to be expressed with their associated variability. *Confidence* refers to the probability the estimated outcome will fall within some level of *precision*. Statement of precision without a statement of *confidence* proves misleading, as evaluation may yield extremely high precision with low confidence or vice versa. For example, after metering a sample of impacted equipment, one may estimate average savings as 1,000 kWh. This is an *estimate* of the *true average* savings. Further, one may able to state the true average is within ±1 percent of the estimate (precision), but only be 30 percent confident that is the case. Alternatively, one may be 99 percent confident the true average savings are within ±50 percent of the estimate of 1,000 kWh.

If the estimated outcomes are large relative to the variation, they tend to be statistically significant. On the other hand, if the amount of variability is large relative to the estimated outcome, one is unable to discern if observed values are real or simply random. In other words, when variability is large, it may lead to precision levels that are too large (e.g., more than ±100 percent) for observed estimates (e.g., estimated savings) to be meaningful. In an extreme example, if the observed average is 1,000 kWh and the associated precision is ±150 percent, true average savings are somewhere between negative 500 kWh (which means the measure actually caused consumption to increase) and 1,500 kWh.

To formalize these relationships, evaluators use a test called the *t* statistic. The *t* statistic is a measure of a statistical estimate's reliability. When the parameter estimate, such as the mean kWh savings, is small relative to its associated variability, the t statistic value is low. In energy efficiency evaluations it is common to use a 95 percent level of confidence, for which the critical value of t is 1.96. If the t statistic is less than 1.96, the evaluator concludes that the estimated value (e.g., mean kWh savings) is not reliable.

Confidence intervals are a convenient way of expressing the potential random sampling error for an estimate. Confidence intervals are calculated by multiplying the estimated standard error by a value based on the t statistic and adding or subtracting this number from the estimate. For example, once average savings are estimated, true average savings are bracketed in the following confidence interval:

estimated average savings–t(SE $_{savings}$)≤true average savings≤estimated average savings +t(SE $_{savings}$)

The rule of thumb is to use a value of 2 times the standard error for calculating a 95 percent confidence. Table D.1 summarizes the statistical terms useful for in assessing uncertainty. (The table provides an easy reference, not a guide for computations.)

For example, assume that 12 monthly energy bills total 48,000 kWh. Estimated average annual consumption is:

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{48{,}000}{12} = 4{,}000$$

The *variance* is:

$$S^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1} = 4{,}488{,}417 \text{ kWh}^2$$

The *standard deviation* is:

$$s = \sqrt{S^2} = \sqrt{4{,}488{,}417} = 2{,}118 \ kHw.$$

The *standard error* is:

$$SE = \frac{s}{\sqrt{n}} = \frac{2{,}118}{\sqrt{12}} = 611 \ kHw$$

Thus, at a 95 percent confidence level, the *absolute precision* is approximately:

$$t \times SE = 2 \times 611 = 1{,}222 \ kWh$$

At a 95 percent confidence level, the *relative precision* is:

$$\frac{t \times SE}{estimate} = \frac{1{,}222}{4{,}000} = 30\%$$

Table D-1. Summary of Statistical Terms

Mean ($\bar{Y}$)	The mean is determined by adding up individual data points and dividing by the total number of these data points.	$\bar{Y} = \frac{\sum Y_i}{n}$
Variance (S^2)	The extent to which observed values differ from each other. Variance is found by averaging the squares of individual deviations from the mean. Deviations from the mean are squared simply to eliminate negative values.	$S^2 = \frac{\sum Y_i (Y_i - \bar{Y})^2}{n-1}$
Standard Deviation (s)	This is simply the square root of the variance. It brings the variability measure back to the units of the data (e.g., while variance units are in kWh2, the standard deviation units are kWh).	$s = \sqrt{S^2}$
Standard Error (SE)	The standard deviation divided by the square root of the total number of observations. SE is the measure of variability used in assessing precision and confidence for the true value of the estimate.	$SE = \frac{s}{\sqrt{n}}$
Coefficient of Variance (cv)	Defined as the standard deviation of the readings divided by the mean, this is used in estimating sample sizes.	$cv = \frac{s}{\bar{Y}}$
Absolute Precision	Computed from standard error using a t value.	$t * SE$
Relative Precision	The absolute precision divided by the estimate.	$\frac{t * SE}{Estimate}$