KANSAS
CORPORATION COMMISSION

Kathleen Sebelius, Governor
Thomas E. Wright, Chairman
Michael C. Moffet, Commissioner
Joseph F. Harkins, Commissioner

Evaluation, Measurement and Verification: A Framing Document

Kansas Corporation Commission
Workshop on Energy Efficiency
March 25 and 26, 2008

1500 SW Arrowhead Road, Topeka, KS 66604-4027 • (785) 271-3100 • Fax: (785) 271-3354 • http://kcc.ks.gov/

Evaluation, Measurement and Verification: A Framing Document

Kansas Corporation Commission
Workshop on Energy Efficiency

Objectives:

- Enable and promote sound regulatory oversight over the costs and savings associated with energy efficiency (and demand response) programs.
- Enable and promote continuous improvement in programs, both to correct program elements that can be improved, and to adapt programs as markets and technology change.
- Produce reliable estimates of demand and energy reductions due to programs for planning purposes, as well as estimates of emissions reductions, if needed.
- Justify incentives, if any.

Backround Document Included Here:
Model Energy Efficiency Program Impact Evaluation Guide produced by the National Action Plan for Energy Efficiency.[1] The Executive Summary is included here.

Further Considerations and Highlights:

Evaluation Budgets:
Some states have evolved to a practice of spending that fits the task. In rough terms and on average, this amounts to 5% of the total budget for energy efficiency. In one case, an arbitrary limit of 3% is applied. For a task so important, the practice leaving the determination of the amount applied to evaluation to Commission discretion is wise. As a side note, a similar comment can apply to program administration. Measuring and limiting inputs, like the cost of administration, can feel fiscally responsible, but this may not optimize outputs, energy savings.

Precision and Accuracy:
Evaluation manages an enormous quantity of information. Accuracy is important, since overall performance in terms of saved energy and capacity is critical for program value. Accuracy should not be confused with precision. Over the scope of a population, an accurate assessment of savings from some programs can be achieved indirectly, through statistical means, or by assuming or deeming that an average amount of savings occurs each time a customer uses the amount applied to evaluation program. In such cases, the precise amount of savings from a specific customer is never counted, except in statistical sampling or when testing the deemed values. Precision comes from metering savings. While this practice works for larger and distinctive or unique customers, it is not practical for mass market customers. Each state has to come to some conclusion about whether to permit indirect means of measurement, and if so, what ways will be permitted.

Who hires and fires the evaluator?:
Because a great deal rides on the results of energy efficiency evaluation, there is reasonable discussion about how this function should be managed. California has taken the view that the

[1] http://www.epa.gov/cleanenergy/energy-programs/napee/resources/guides.html#guide5 (March 15, 2008)

state should manage the evaluation process, hiring and supervising evaluation contractors. Further, evaluators in California cannot also work in program design or implementation for California utilities or others who administer programs. This separation draws a bright line around the evaluation process. California, with its size, experience with energy efficiency (this stand was taken just a few years ago) and ample consultant population, may be able to draw this distinction in ways that smaller states with fewer local consultants would not. At the other end of the spectrum, some utilities are permitted to do evaluation work with employees, though there is usually some administrative separation between program implementers and evaluators.

Commissions generally assess their situation (i.e. utility attitude toward efficiency, degree to which incentives are connected to performance, commission and utility staffing) to find the right balance of accountability and manageability to get the right evaluation program standard practice.

Measurement in it proper place:
There is an appropriate motivation to design programs to make measurement easy, but this can go too far. For example, one way to measure savings is to give and count rebates. A problem can emerge if experience indicates that a different program design will be more effective at producing savings. Adherence to the rebate device can count too heavily in this decision.

Reporting:
Efficiency portfolio managers should report results each year. This practice assures a regular check in on the quality and emphasis of the programs, and assures that the administrator and others can consider how the programs can be improved, and how the emphasis of the portfolio can change to take advantage of new opportunities in the market. Commissions sometimes consider putting utilities on different schedules so review can be spread throughout the year, rather then occurring in a bubble between March and May every year.

Experience in Evaluation is a Strength:
As the National Action Plan for Energy Efficiency Guide to Evaluation indicates, there is a significant body of experience concerning evaluation of energy efficiency. Early stage bugs have been worked out of the system in most states. Methods for estimating measure persistence, net savings, spillover effects (savings that customer acquire from energy efficiency investments they make on their own after being influenced by a program) and other factors are well-developed. Commissions should have confidence that accurate assessments of energy efficiency savings can occur.

Assessing the savings from market transformation programs – efforts designed to change the overall workings of the energy efficiency marketplace, but which may produce little or no direct savings in the near term – remains more of a work in progress. This leads some states (Efficiency Vermont is an example) at the outset to emphasize savings-oriented programs at the start to allow the programs to post tangible gains at a modest cost, leaving market transformation to develop later. On the other hand, Arkansas is featuring a market transformation program among its Quick Start programs.

5: Calculating Net Energy and Demand Savings

Chapter 5 defines net savings and describes the four key factors that differentiate net and gross savings: free ridership, spillover effects, rebound effects, and electricity transmission and distribution losses. The chapter then provides a detailed description of several approaches for determining net savings, including self-reporting surveys, econometric models, and stipulated net-to-gross ratios. A brief discussion of the criteria for selecting an appropriate net savings evaluation approach is also provided.

5.1 Importance of Net Savings

To keep program benefits from being under- or overstated, it is important to understand and properly reflect the influences of both energy savings and emission avoidance programs. These net savings are the savings "net" of what would have occurred in the absence of the program. Generally speaking, net savings are of most interest for regulated government and utility programs. In these cases, the responsible party (for example, a city council or utility regulator) wants to know if the use of public or ratepayer funded programs are actually having an influence. That is, are the programs of interest providing incremental benefits, or do the benefits result from some other influences? For example, the environmental benefits of energy efficiency programs are usually considered valid only if they are additional to naturally occurring efficiency activities (that is, based on net savings). In contrast, private sector energy efficiency programs such as performance contracts are a case where gross energy savings are the primary concern.

The following sections describe factors that differentiate net and gross impacts and approaches for calculating NTGRs. It is important to understand, though, that calculating net energy and demand savings can be more of an art than a science. Essentially, one is attempting to separate out the influence of a particular energy efficiency program (or portfolio) from all the other influences that determine participant and non-participant behavior and decisions. With the increasing "push" for energy efficiency by utilities and government at the local, state, and national level and by private groups and large companies, it can be quite difficult to separate out how one particular program among all this activity influences the decision of whether, when, and to what degree to adopt efficiency actions.

5.2 Factors That Account for Differences Between Net and Gross Savings

The three primary factors that differentiate gross and net savings are free ridership, spillover, and rebound. In addition, transmission and distribution losses can also be considered under a NTGR calculation for programs that save electricity from grid-connected power plants. The decision about which of these to include in an NTGR analysis is determined by the objectives of the evaluation. Free ridership is typically the most commonly evaluated NTGR factor, followed by spillover and then rebound analyses.

- **Free ridership.** Free riders are program participants who would have implemented the program measure or practice in the absence of the program. The program can also affect when a participant implements an efficiency measure (e.g., because of the program a participant installs the equipment sooner than he or she otherwise would have), the level of efficiency of the efficient equipment installed (e.g., a participant says he or she would have installed the same efficient equipment without the program), and the number of units of efficiency equipment installed. Different levels of free ridership introduce the concept of partial or deferred free riders. The subjectivity surrounding free ridership is a significant component of net energy and demand savings uncertainty.

Free Riders

There are three categories of free riders:

- **Total free rider**—would have installed the same energy efficiency measures at the same time whether or not the program existed.
- **Partial or deferred free rider**—would have installed less-efficient (but still more efficient than baseline) measures or would have installed the same energy efficiency measure but at a later time and would have installed fewer of the energy efficiency products.
- **Non-free rider**—would not have installed the baseline energy efficiency measure without the influence of the program.

It should be noted that a participant's free ridership status can vary from one measure to the next and over time.

- **Spillover effects.** Spillover occurs when there are reductions in energy consumption or demand caused by the presence of the energy efficiency program, but which the program does not directly influence. Customer behavioral changes stemming from participation in programs are a positive program spillover, increasing the program effect. These effects could result from (a) additional energy efficiency actions that program participants take outside the program as a result of having participated; (b) changes in the array of energy-using equipment that manufacturers, dealers, and contractors offer all customers (and they purchase) as a result of program availability; (c) changes in specification practices employed by architects and engineers; and (d) changes in the energy use of non-participants as a result of utility programs, whether direct (e.g., utility program advertising) or indirect (e.g., stocking practices such as (b) above, or changes in consumer buying habits). The term "free driver" is used to describe a non-participant who has adopted a particular efficiency measure or practice as a result of a utility program.

The analysis of spillover and free ridership is complicated by "market noise." When a market is filled with many implementers offering similar programs under different names, with different incentive structures and marketing methods, it is difficult to estimate any particular program's influence. Identification of non-participants may also be difficult, since customers may not be able to discern between the various programs operating in the marketplace and may not accurately recall how programs may have influenced their decision processes or even remember the program in which they participated.

- **Rebound effect.** Rebound is a change in energy-using behavior that increases the level of service and results from an energy efficiency action. The most common form is "take back," which can occur if consumers increase energy use as a result of a new device's improved efficiency. For example, homeowners may use more air-conditioning with their new efficient air-conditioner because it is cheaper to run than their old air-conditioner. Another example is when insulation is installed for a low-income household and the homeowner can turn the thermostat up to a more comfortable temperature. However, there is a non-energy benefit here associated with increased comfort, health, and safety that some would argue should be considered a co-benefit.

Rebound effect is part of the general concept of how customer behavior affects technology usage and, thus, efficiency performance. For example, installation of occupancy sensors in small independent hotels would not save energy if hotel staff were already adjusting HVAC manually as part of their ordinary maintenance. In another example, an Energy Management System could be overridden by management decisions. Behavioral issues such as these are becoming of increasing interest in advanced energy efficiency programs.

- **Electricity transmission and distribution losses.** When an efficiency project reduces electricity consumption at a facility, the amount of electricity that no longer has to be generated at a power plant is actually greater than the onsite reduction. This is because of electricity transmission and distribution (T&D) losses between the sites and the power plants. Published electricity grid emission factors do not usually include T&D losses and most energy savings evaluations only report onsite energy savings. Therefore an evaluator needs to decide whether to include T&D losses in their net savings calculation.

T&D losses can range from negligible for a high-voltage customer located close to a power plant to over 10% for smaller customers located far from power plants. In addition, higher T&D losses are inevitable during on-peak hours. Thus, some jurisdictions have calculated on-peak, off-peak, and seasonal T&D loss factors.

If a T&D loss factor is being considered, it is best to adopt one factor (or perhaps two, one for on-peak and one for off-peak) for the entire grid and not attempt to be too fine-grained. Two options for quantifying T&D losses are (a) assuming a simple percentage adder for source savings and (b) not including T&D losses directly, but considering them a counterweight to uncertainty in the site savings calculation. The adder could be a value calculated for the specific T&D network in question. Potential sources of such data are local regulatory authorities, local utilities, and the regional independent system operator (ISO).

EPA's Conservation Verification Protocol (EPA, 1995) for the Acid Rain Program suggests the following default values for T&D losses, as a proportional adder to onsite energy savings:

- T&D savings for residential and commercial customers—7 percent
- T&D savings for industrial customers—3.5 percent

This consideration of T&D issues is often part of a calculation to determine "source" energy (fuel) savings (i.e., how much fuel is not consumed in a power plant because of the end-use efficiency activity).

Source fuel savings are calculated by considering both T&D losses and power plant fuel efficiencies. *It should also be noted that T&D losses and source energy savings calculations are often considered in the gross energy savings calculation instead of the net energy savings calculation*. In either case, savings should be reported with an indication of whether they include T&D losses and are based on source energy or end-use energy.

Other influences (in addition to free ridership, spillover, rebound, and T&D losses) that can determine net versus gross savings include:

- The state of the economy (recession, recovery, economic growth).
- Energy prices.
- Changes in facility operations (e.g., office building or hotel occupancy rates, changes in product lines or number of operating shifts in factors, or changes in thermostat settings or number of people living in homes). These are typically addressed in the gross savings analyses.

5.3 Approaches for Determining Net Savings

The following discussion presents the four approaches for determining the NTGR:

- **Self-reporting surveys.** Information is reported by participants and non-participants without independent verification or review.
- **Enhanced self-reporting surveys.** The self-reporting surveys are combined with interviews and documentation review and analysis.
- **Econometric methods.** Statistical models are used to compare participant and non-participant energy and demand patterns.
- **Stipulated net-to-gross ratios.** Ratios that are multiplied by the gross savings to obtain an estimate

of net savings and are based on historical studies of similar programs.

With respect to program size and scale, the two survey methods can be used with any program regardless of the number of participants. The third approach can only be used with programs with large numbers of participants because the models need large amounts of data to provide reliable results. The fourth approach can be used any time there is sufficient data to support a stipulated value.

In terms of timing, an NTGR analysis can be integrated into the gross impact analysis if the large-scale data analysis approach is used. With other gross impact analysis approaches, the NTGR is calculated independently, perhaps covering a longer period of time to more fully cover spillover and rebound effects. However, as with gross impact analysis, some of the approaches can be costly and evaluation resources can be limited. Accordingly, it is acceptable to perform NTGR analyses less frequently than the gross savings impact evaluation—perhaps every few years—as long as the market influences and participants' behavior are relatively consistent.

In terms of accuracy requirements, while econometric modeling can include tests for bias and precision and appropriate sample sizes can be determined, it is virtually impossible to define a precision target and a statistically valid sample size for the two self-reporting survey approaches. This challenge in surveying comes from the nature of collecting both qualitative and quantitative data from various participants and non-participants involved in the decision to install energy efficiency measures. In this case, evaluators attempt to survey all participants or intuitively select survey sample sizes.

The other uncertainty challenge in surveying is the subjective nature of assigning NTGRs to each participant. A participant is clearly a free rider if he or she would have installed the same project even if the program did not exist. Assigning NTGRs to individual participants is more complicated in cases where a participant *might* have installed the project, or would have installed it in two years if not for the program.

When non-participants are included in the NTGR analysis, care must be taken in selecting the appropriate comparison group. There is no single rule about what constitutes an appropriate comparison group, since the selection of the group depends on such factors as type of market transaction, survey methodology, and comparison purpose. The proposed non-participant comparison group and the criteria used in selecting this group should be discussed in the evaluation plan.

The following subsections briefly discuss the four approaches. (More information, specific to energy efficiency NTGR evaluations, can be found in CPUC, 2004.)

5.3.1 Self-Reporting Surveys

Survey-based stated intentions, or "self-reports," are a way to estimate free ridership by asking participants a series of questions on what they would have done in the absence of the program. Spillover estimates are developed and free ridership estimates are enhanced by non-participant surveys.

Surveys can be surprisingly complex to design and administer. They rely on respondent selection methods, survey instrument design, question wording, and implementation method to develop reliable results. One of the elements that should be addressed in surveys is self-selection bias. Self-selection bias is possible whenever the group being studied has any form of control over whether to participate: for example, people with strong opinions or substantial knowledge may be more willing to spend time answering a survey than those who do not. Self-selection bias is related to sample selection bias and can skew the results of an NTGR analysis that is not very well planned, funded, or executed.

Generally, the best use of self-reporting surveys has involved asking a series of questions with each question allowing a scale of responses. Surveys are either hard copy or Web-based instruments that are filled out by the interviewee, or perhaps conducted by phone with a professional surveyor (usually someone unfamiliar with energy efficiency). A typical initial question asked of participants is, "If the program had not existed, would you have installed the same equipment?" For a response, participants might choose between "definitely would have," "probably would have," "probably would not have," and "definitely would not have." This use of a scale, rather than a yes/no response, is thought to allow greater apparent confidence and precision in the estimate.

For free ridership, each of the responses is assigned a probability to determine the expected net savings. These estimates are then combined (additively or multiplicatively) into an individual participant free rider estimate. The participant estimates are subsequently averaged (or assigned a weighted average based on expected savings) to calculate the overall free ridership estimate. Similarly, non-participant responses are used to adjust a free ridership estimate and/or calculate spillover estimates.

Table 5-1 provides an example of a probability matrix used to determine a free ridership score. Note that the only 100 percent free ridership score is attained if a measure was already on order or installed prior to participation in the program. This approach was used in a process and impact evaluation of the Southern California Edison IDEEA program and an impact evaluation of the Energy Trust of Oregon's commercial and industrial programs.[1] (Note that the content of this table is intended only to illustrate the basic concepts.)

The survey approach is the most straightforward way to estimate free ridership and spillover. It is also the lowest-cost approach. It does, however, have its disadvantages in potential bias and with overall accuracy. For example, typical responses such as "don't know," missing data, and inconsistent answers are very hard to address without additional data collection. While there are ways to improve survey quality (e.g., using techniques like adding consistency check questions and adjusting the individual's estimate accordingly), the accuracy of simple self-reports is typically marginal.

5.3.2 Enhanced Self-Reporting Surveys

To improve the quality of NTGRs drawn from self-reported survey responses, the evaluation can rely on multiple data sources for the decision to install or adopt energy efficiency measures or practices. Some common additional data sources and techniques are:

- **Personal surveys.** Conducting in-person surveys is probably the best way to qualitatively improve the quality of self-surveys. Key participants in the decision to install efficiency measures can help determine the level of influence of the program on participants and non-participants. For commercial and government facilities, potential interviewees include managers, engineers, and facilities staff. Contractors, design engineers, and product manufacturers, distributors, and retailers can also provide information on the

Table 5-1. Example Free Rider Probability Assessment

Free-Ridership Score	Already Ordered or Installed	Would Have Installed Without Program	Same Efficiency	Would have Installed All of the Measures	Planning to Install Soon	Already in Budget
100%	Yes	Yes	—	—	—	—
0%	No	No	—	—	—	—
0%	No	Yes	No	—	—	—
50%	No	Yes	Yes	Yes	Yes	Yes
25%	No	Yes	Yes	Yes	No	Yes
25%	No	Yes	Yes	Yes	Yes	No
0%	No	Yes	Yes	Yes	No	No
25%	No	Yes	Yes	No	Yes	Yes
12.5%	No	Yes	Yes	No	No	Yes
12.5%	No	Yes	Yes	No	Yes	No
0%	No	Yes	Yes	No	No	No

Provided by Sami Khawaja of Quantec, LLC.

influences and motivations that determine the role of energy efficiency programs in the decision-making process. When working with professionals involved in the efficiency measure installation, individuals familiar with the program and projects should conduct the interviews. The interviewer should attempt to eliminate or at least minimize any bias they may have.

- **Project analysis.** This consists of two general types of reviews. The first is an analysis of the barriers to project installation and how the project addresses these barriers. The most common barrier is financial (project costs), so the common analysis is calculation of a project's simple payback. For example, if the project has a very short payback period without any program-provided benefits, then it may be considered as more likely to have been installed with or without the program.[2] The other type of analysis is reviewing any documentation the participant may have of the decision to proceed with the project. Such documentation may include internal memos or feasibility studies and can indicate the basis of the decision to proceed.

- **Non-specific market data collection.** Through the review of other information resources prepared for similar programs, the range of appropriate NTGRs can be estimated. Such resources might include analyses of market sales and shipping patterns, studies of decisions by participants and non-participants in similar programs, and market assessment, potential, or effects studies. Market sales methods rely on aggregate data on total sales of a particular technology in a given jurisdiction. They compare this sales volume with a baseline estimate of the volume that would have been sold in the absence of the program. The accuracy of these methods depends on the completeness and accuracy of the sales data, as well as the validity of the baseline estimate.

All or some of these three data sources can be combined with the written or Web-based participant and non-participant self-surveys to triangulate on an estimate of the free ridership, spillover, and rebound rates for that program.

Net-to-Gross Ratio Calculation Using Equipment Sales Data

In 1992 Baltimore Gas and Electric (BGE) offered a number of conservation programs, including a residential HVAC program. This program was designed to give consumers who were in the market to replace their HVAC systems incentives to choose a more energy-efficient heat pump or central air conditioner. BGE conducted an impact evaluation including a net-to-gross analysis designed to quantify the portion of energy-efficient HVAC purchases that could be attributed to BGE's program. Several sources of data were used:

- A survey of participants in BGE's residential HVAC program.
- Two surveys of customers who did not participate in BGE's residential HVAC programs.
- A survey of HVAC contractors who reported their sales of HVAC equipment by SEER (seasonal energy efficiency ratio).
- Data from the Air Conditioning and Refrigeration Institute that provided SEER levels for central air conditioners and heat pumps on an annual basis from 1981 through 1991.

These data provide a range of NTGRs from 0.74 to 0.92. An integrated approach provided what BGE considered the most reliable estimate:

Net-to-gross ratio = Net increase in purchases of qualifying equipment due to the program divided by the number of units sold under the program in 1992

= (28,300 – 18,700) ÷ 10,400
= 0.92

Thus, BGE concluded that an initial NTGR of 0.90 was appropriate.

Case study provided by Baltimore Gas and Electric.

5.3.3 Econometric Models

Econometric models, in this context, are mathematical tools that apply quantitative or statistical methods to the analysis of NTGRs. Econometric methods are sometimes considered the most accurate approach to calculating NTGRs when there are enough participants and truly comparable non-participants, and when the program is large enough to justify the cost of such analyses. The econometric models are closely related to, and can be the same models as, those described in Section 4.3 for calculating gross energy savings.

Various econometric methods have been used, with varying advantages and disadvantages. The models use energy (and demand) data from participants and non-participants over the same period to estimate the difference between gross savings (participant savings) and simple net savings (participant savings minus non-participant savings). The models differ in their mathematical and statistical calculation methods, but also in how they address complicating factors of bias that differentiate true NTGRs from simple comparisons of participant and non-participant savings. One particular element of surveying that econometric models attempt to address is self-selection bias.

5.3.4 Stipulated Net-to-Gross Ratio

This fourth approach, although not a calculation approach, is often used. NTGRs are stipulated in some jurisdictions when the expense of conducting NTGR analyses and the uncertainty of the potential results are considered significant barriers. In such a situation, a regulatory body sets the value, which is typically in the 80 to 95 percent range. Sources of stipulated NTGRs should be similar evaluations of other programs or, possibly, public utility commissions' requirements. Stipulated NTGRs should be updated periodically based on evaluations and review of other programs' calculated NTGRs.

5.4 Selecting a Net Savings Evaluation Approach

As mentioned in Chapter 4, selection of an evaluation approach is tied to the objectives of the program being evaluated, the scale of the program, the evaluation budget and resources, and specific aspects of the measures and participants in the program.

Another criterion—probably the most important—is cost. All four approaches can be used with any type of efficiency program, with the possible exception that the econometric modeling requires a program with a large number of participants. The lowest-budget approach is stipulated NTGR, followed by self-reporting surveys and enhanced surveys, and then econometric modeling (which incorporates the surveying activities). One option for keeping costs down while using the more sophisticated approaches is to conduct an NTGR analysis every few years and stipulate NTGRs for the intervening years.

5.5 Notes

1. Provided courtesy of Quantec, LLC.

2. Note that the need to decide when a consumer would have installed an energy project, based on the economic payback associated with a project, is an example of the subjective nature of free ridership. The choice of a specific payback period—2, 3, 4, etc., years—to define who is and who is not a free rider also has a subjective nature.