

**Methodological Framework for Using the Self-
Report Approach to Estimating Net-to-Gross
Ratios for Nonresidential Customers**

**Prepared for the Energy Division, California Public Utilities
Commission**

By

The Nonresidential Net-To-Gross Ratio Working Group

Final Version

TABLE OF CONTENTS

1. OVERVIEW OF THE LARGE NONRESIDENTIAL FREE RIDERSHIP APPROACH	1
2. BASIS FOR SRA IN SOCIAL SCIENCE LITERATURE	2
3. FREE RIDERSHIP ANALYSIS BY PROJECT TYPE	2
4. SOURCES OF INFORMATION ON FREE RIDERSHIP	2
5. NTGR FRAMEWORK	5
5.1. NTGR Questions and Scoring Algorithm	6
5.1.1. Timing and Selection Score 7	
5.1.2. Program Influence Score 8	
5.1.3. No-Program Score 8	
5.1.4. The Core NTGR 9	
5.2. Data Analysis and Integration	9
5.3. Accounting for Partial Free Ridership	13
6. NTGR INTERVIEW PROCESS	15
7. DATA INTEGRATION RULES	ERROR! BOOKMARK NOT DEFINED.
8. COMPLIANCE WITH SELF-REPORT GUIDELINES	16

Appendix A: References

Appendix B: Net-to-Gross Questions and Uses of Data by Level of NTGR Analysis

Appendix C: NTGR Scoring Algorithm and Example

Appendix D: Demonstration of Compliance with the CPUC/ED Guidelines for Estimating Net-to-Gross Ratios Using the Self-Report Approach

Acknowledgments

As part of the evaluation of the 2006-08 energy efficiency programs designed and implemented by the four investor-owned utilities (Pacific Gas & Electric Company, Southern California Edison Company, Southern California Gas Company, and San Diego Gas and Electric Company) and third parties, the Energy Division of the California Public Utilities Commission (CPUC) formed a nonresidential net-to-gross ratio working group that was composed of experienced evaluation professionals. The main purpose of this group was to develop a standard methodological framework, including decision rules, for integrating in a systematic and consistent manner the findings from both quantitative and qualitative information in estimating net-to-gross ratios. The working group, listed alphabetically, was composed of the following evaluation professionals:

- Michael Baker, SBW Consulting
- Fred Coito, KEMA
- Kevin Cooney, Summit Blue Consulting
- Tim Drew, Energy Division, CPUC
- Jennifer Fagan, Itron, Inc.
- Miriam Goldberg, KEMA
- Nick Hall, TecMarket Works
- Kay Hardy, Energy Division, CPUC
- Ken Keating
- John Reed, Innovologie LLC
- Richard Ridge, Ridge & Associates
- Mike Rufo, Itron, Inc.
- Eric Swan, KEMA (formerly of RLW Analytics, Inc.)
- Christina Torok, Itron, Inc.
- Philippus Willems, PWP, Inc.

A public webinar was conducted to obtain feedback from the four investor-owned utilities and other interested stakeholders. The questionnaire was then pre-tested and, based on the pre-test results, finalized in November 2008.

1. OVERVIEW OF THE LARGE NONRESIDENTIAL FREE RIDERSHIP APPROACH

The methodology described in this section was developed to address the unique needs of Large Nonresidential customer projects developed through energy efficiency programs offered by the four California investor-owned utilities and third-parties. This method relies exclusively on the Self-Report Approach (SRA) to estimate project and program-level Net-to-Gross Ratios (NTGRs), since other available methods and research designs are generally not feasible for large nonresidential customer programs. This methodology provides a standard framework, including decision rules, for integrating findings from both quantitative and qualitative information in the calculation of the net-to-gross ratio in a systematic and consistent manner. This approach is designed to fully comply with the *California Energy Efficiency Evaluation: Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals* (Protocols) and the *Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approaches* (Guidelines), as demonstrated in Appendix D.

This approach preserves the most important elements of the approaches previously used to estimate the NTGRs in large nonresidential customer programs¹. However, it also incorporates several enhancements that are designed to improve upon that approach, for example:

- The method introduces a 0 to 10 scoring system for key questions used to estimate the NTGR, rather than using fixed categories that were assigned weights (as was done previously).
- The method asks respondents to jointly consider and rate the importance of the many likely events or factors that may have influenced their energy efficiency decision making, rather than focusing narrowly on only their rating of the program's importance. This question structure more accurately reflects the complex nature of the real-world decision making and should help to ensure that all non-program influences are reflected in the NTGR assessment in addition to program influences.

It is important to note that the NTGR approach described in this document is a general framework, designed to address all large nonresidential programs. In order to implement this approach on a program-specific basis, it might need to be somewhat customized to reflect the unique nature of the individual programs.

¹ Such as, for example, the NTGR method used to evaluate NTGRs for the California Standard Performance Contracting Program.

2. BASIS FOR SRA IN SOCIAL SCIENCE LITERATURE

The social sciences literature provides strong support for use of the methods used in the SRA to assess program influence. As the *Guidelines* notes,

More specifically, the SRA is a mixed method approach that involves asking one or more key participant decision-makers a series of structured and open-ended questions about whether they would have installed the same EE equipment in the absence of the program as well as questions that attempt to rule out rival explanations for the installation (Weiss, 1972; Scriven, 1976; Shadish, 1991; Wholey et al., 1994; Yin, 1994; Mohr, 1995). In the simplest case (e.g., residential customers), the SRA is based primarily on quantitative data while in more complex cases the SRA is strengthened by the inclusion of additional quantitative and qualitative data which can include, among others, in-depth, open-ended interviews, direct observation, and review of program records. Many evaluators believe that additional qualitative data regarding the economics of the customer's decision and the decision process itself can be very useful in supporting or modifying quantitatively-based results (Britan, 1978; Weiss and Rein, 1972; Patton, 1987; Tashakkori and Teddlie, 1998).²

More details regarding the philosophical and methodological underpinnings of this approach are in Ridge, Willems and Fagan (2009), Ridge, Willems, Fagan and Randazzo (2009) and Megdal, Patil, Gregoire, Meissner, and Parlin (2009). In addition to these two articles, Appendix A provides an extensive listing of references in the social sciences literature regarding the methods employed in the SRA.

3. FREE RIDERSHIP ANALYSIS BY PROJECT TYPE

There are three levels of free-ridership analysis. The most detailed level of analysis, the **Standard – Very Large Project** NTGR, is applied to the largest and most complex projects (representing 10 to 20% of the total) with the greatest expected levels of gross savings.³ The **Standard** NTGR, involving a somewhat less detailed level of analysis, is applied to projects with moderately high levels of gross savings. The least detailed analysis, the **Basic** NTGR, is applied to all remaining projects. Evaluators must exercise their own discretion as to what the appropriate thresholds should be for each of these three levels.

4. SOURCES OF INFORMATION ON FREE RIDERSHIP

There are five sources of free-ridership information in this study. Each level of analysis relies on information from one or more of these sources. These sources are described below.

² *Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approaches*, October 15, 2007, pg. 3.

³ Note that we do not refer to an Enhanced level of analysis, since this is defined by the Protocols to involve the application of two separate analysis approaches, such as billing analysis or discrete choice modeling.

1. **Program Files.** As described in previous sections of this report, programs often maintain a paper file for each paid application. These can contain various pieces of information which are relevant to the analysis of free-ridership, such as letters written by the utility's customer representatives that document what the customer had planned to do in the absence of the rebate and explain the customer's motivation for implementing the efficiency measure. Information on the measure payback with and without the rebate may also be available.

2. **Decision-Maker Surveys.** When a site is recruited, one must also determine who was involved in the decision-making process which led to the implementation of measures under the program. They are asked to complete a Decision Maker survey. This survey obtains highly structured responses concerning the probability that the customer would have implemented the same measure in the absence of the program. First, participants are asked about the timing of their program awareness relative to their decision to purchase or implement the energy efficiency measure. Next, they are asked to rate the importance of the program versus non-program influences in their decision making. Third, they are asked to rate the significance of various factors and events that may have led to their decision to implement the energy efficiency measure at the time that they did. These include:
 - the age or condition of the equipment,
 - information from a feasibility study or facility audit
 - the availability of an incentive or endorsement through the program
 - a recommendation from an equipment supplier, auditor or consulting engineer
 - their previous experience with the program or measure,
 - information from a program-sponsored training course or marketing materials provided by the program
 - the measure being included as part of a major remodeling project
 - a recommendation from program staff, a program vendor, or a utility representative
 - a standard business practice
 - an internal business procedure or policy
 - stated concerns about global warming or the environment
 - a stated desire to achieve energy independence.

In addition, the survey obtains a description of what the customer would have done in the absence of the program, beginning with whether the implementation was an early replacement action. If it was not, the decision maker is asked to provide a description of what equipment would have been implemented in the absence of the program, including both the efficiency level and quantities of these alternative measures. This is used to adjust the gross engineering savings estimate for partial free ridership, as discussed in Section 5.2.

This survey contains a core set of questions for **Basic** NTGR sites, and several supplemental questions for both **Standard** and **Standard – Very Large** NTGR

sites For example, if a Standard or Standard-Very Large respondent indicates that a financial calculation entered highly into their decision, they are asked additional questions about their *financial criteria* for investments and their rationale for the current project in light of them. Similarly, if they respond that a *corporate policy* was a primary consideration in their decision, they are asked a series of questions about the specific policy that led to their adoption of the installed measure. If they indicate the installation was a *standard practice*, there are supplemental questions to understand the origin and evolution of that standard practice within their organization. These questions are intended to provide a deeper understanding of the decision making process and the likely level of program influence versus these internal policies and procedures. Responses to these questions also serve as a basis for consistency checks to investigate conflicting answers regarding the relative importance of the program and other elements in influencing the decision. In addition, **Standard – Very Large** sites may receive additional detailed probing on various aspects of their installation decision based on industry- or technology-specific issues, as determined by review of other information sources. For Standard-Very Large sites all these data are used to construct an internally consistent “story” that supports the NTGR calculated based on the overall information given.

3. **Vendor Surveys.** A Vendor Survey is completed for all **Standard** and **Standard-Very Large** NTGR sites that utilized vendors, and for **Basic** NTGR sites that indicate a high level of vendor influence in the decision to implement the energy efficient measure. For those sites that indicate the vendor was very influential in decision making, the vendor survey results enter directly into the NTGR scoring. The vendor survey findings are also be used to corroborate Decision Maker findings, particularly with respect to the vendor’s specific role and degree of influence on the decision to implement the energy efficient measure. Vendors are queried on the program’s significance in their decision to recommend the energy efficient measures, and on their likelihood to have recommended the same measure in the absence of the program. Generally, the vendors contacted as part of this study are contractors, design engineers, distributors, and installers.
4. **Utility and Program Staff Interviews.** For the Standard and Standard-Very Large NTGR analyses, interviews with utility staff and program staff are also conducted. These interviews are designed to gather information on the historical background of the customer’s decision to install the efficient equipment, the role of the utility and program staff in this decision, and the name and contact information of vendors who were involved in the specification and installation of the equipment.
5. **Other information.** For **Standard – Very Large Project** NTGR sites, secondary research of other pertinent data sources is performed. For example, this could include a review of standard and best practices through industry associations, industry experts, and information from secondary sources (such as the U.S. Department of Energy's Industrial Technologies Program, Best Practices website URL, <http://www1.eere.energy.gov/industry/bestpractices/>). In addition, the Standard- Very Large NTGR analysis calls for interviews with other employees at the participant’s firm, sometimes in other states, and equipment vendor experts

from other states where the rebated equipment is being installed (some without rebates), to provide further input on standard practice within each company.

Table 1 below shows the data sources used in each of the three levels of free-ridership analysis. Although more than one level of analysis may share the same source, the amount of information that is utilized in the analysis may vary. For example, all three levels of analysis obtain core question data from the Decision Maker survey.

Table 1: Information Sources for Three Levels of NTGR Analysis

	Program File	Decision Maker Survey Core Question	Vendor Surveys	Decision Maker Survey Supplemental Questions	Utility & Program Staff Interviews	Other Research Findings
Basic NTGR	√	√	√ ¹		√ ²	
Standard NTGR	√	√	√ ¹	√	√	
Standard NTGR - Very Large Projects	√	√	√ ³	√	√	√

¹Only performed for sites that indicate a vendor influence score (N3d) greater than maximum of the other program element scores (N3b, N3c, N3g, N3h, N3I).

²Only performed for sites that have a utility account representative

³Only performed if significant vendor influence reported or if secondary research indicates the installed measure may be becoming standard practice.

Appendix B provides the full battery of Decision Maker and Vendor survey questions along with notes, for each NTGR level, regarding which questions are asked (denoted by an “X”), and the intended uses of the information in the NTGR analysis. In the case of Basic sites, “TRIGGER” means that a vendor influence score greater than the maximum of other program element scores (N3b, N3c, N3g, N3h, N3I) triggers a vendor survey. In the case of Standard and Standard-Very Large NTGR sites, “TRIGGER” means that a score of 6 or greater triggers a further investigation. A copy of the complete survey forms (with lead-in text and skip patterns) are contained in *Final Large Nonresidential NTGR Survey Instruments.XLS* that is available upon request.

5. NTGR FRAMEWORK

The Self-Report-based Net-to-Gross analysis relies on responses to a series of survey questions that are designed to measure the influence of the program on the participant’s decision to implement program-eligible energy efficiency measure(s). Based on these

responses, a NTGR is derived based on responses to a set of “core” NTGR questions. The NTGR includes the effects of deferred free ridership (i.e., accelerated adoption).

5.1. NTGR Questions and Scoring Algorithm

A self-report NTGR is computed for all NTGR levels using the following approach. Adjustments may be made for **Standard – Very Large** NTGR sites, if the additional information that is collected is inconsistent with information provided through the Decision Maker survey.

The NTGR is calculated as an average of three scores. Each of these scores represents the highest response or the average of several responses given to one or more questions about the decision to install a program measure.

1. A **Timing and Selection** score that reflects the influence of the **most important** of various program and program-related elements in the customer’s decision to select the specific program measure at this time. Program influence through vendor recommendations is also incorporated in this score.
2. A **Program Influence** score that captures the perceived importance of the program (whether rebate, recommendation, training, or other program intervention) relative to non-program factors in the decision to implement the specific measure that was eventually adopted or installed. This score is determined by asking respondents to assign importance values to both the program and most important non-program influences so that the two total 10. The program influence score is adjusted (i.e., divided by 2) if respondents say they had already made their decision to install the specific program qualifying measure before they learned about the program.
3. A **No-Program** score that captures the likelihood of various actions the customer might have taken at this time and in the future if the program had not been available (the counterfactual). This score also accounts for deferred free ridership by incorporating the likelihood that the customer would have installed program-qualifying measures at a later date if the program had not been available.

When there are multiple questions that feed into the scoring algorithm, as is the case for both the **Timing and Selection** and **No-Program** scores, the maximum score is always used. The rationale for using the maximum value is to capture the most important element in the participant’s decision making. Thus, each score is always based on the strongest influence indicated by the respondent. However, high scores that are inconsistent with other previous responses trigger consistency checks and can lead to follow-up questions to clarify and resolve the discrepancy.

The calculation of each of the above scores is discussed below. For each score, the associated questions are presented and the computation of each score is described. For a detailed explanation of the scoring algorithm, including examples, see Appendix C.

5.1.1. Timing and Selection Score

For the Decision Maker, the questions asked are:

I'm going to ask you to rate the importance of the program as well as other factors that might influence your decision to implement [MEASURE.] Think of the degree of importance as being shown on a scale with equally spaced units from 0 to 10, where 0 means not at all important and 10 means very important, so that an importance rating of 8 shows twice as much influence as a rating of 4.

Now, using this 0 to 10 rating scale, where 0 means “Not at all important” and 10 means “Very important,” please rate the importance of each of the following in your decision to implement this specific [MEASURE] at this time.

- Availability of the PROGRAM rebate
- Information provided through a recent feasibility study, energy audit or other types of technical assistance provided through PROGRAM
- Information from PROGRAM training course
- Information from other PROGRAM marketing materials
- Recommendation from a vendor/supplier (If a score of greater than 5 is given, a vendor interview is triggered)

For the Vendor, the questions asked (if the interview is triggered) are:

I'm going to ask you to rate the importance of the [PROGRAM] in influencing your decision to recommend [MEASURE] to [CUSTOMER] and other customers. Think of the degree of importance as being shown on a scale with equally spaced units from 0 to 10, where 0 means not at all important and 10 means very important, so that an importance rating of 8 shows twice as much influence as a rating of 4.

1. Using this 0 to 10 scale where 0 is “Not at all important” and 10 is “Very Important,” how important was the PROGRAM, including incentives as well as program services and information, in influencing your decision to recommend that CUSTOMER install the energy efficiency MEASURE at this time?
2. And using a 0 to 10 likelihood scale, where 0 denotes “not at all likely” and 10 denotes “very likely,” if the PROGRAM, including incentives as well as program services and information, had not been available, what is the likelihood that you would have recommended this specific energy efficiency MEASURE to CUSTOMER?
3. Now, using a 0 to 100 percent scale, in what percent of sales situations did you recommend MEASURE before you learned about the [PROGRAM]?
4. And using the same 0 to 100 percent scale, in what percent of sales situations do you recommend MEASURE now that you have worked with the [PROGRAM]?

5. And, using the same 0 to 10 scale where 0 is “Not at all important” and 10 is “Very important”, how important in your recommendation were:
 - a. Training seminars provided by UTILITY?
 - b. Information provided by the UTILITY website?
 - c. Your firm’s past participation in a rebate or audit program sponsored by UTILITY?

If the Vendor interview is triggered, a score is calculated that captures the highest degree of program influence on the vendor’s recommendation. This score (VMAX) is calculated as the MAXIMUM value of the following:

1. The response to question 1
2. 10 minus the response to question 2
3. The response to question 4 minus the response to question 3, divided by 10
4. The response to question 5a.
5. The response to question 5b.
6. The response to question 5c.

Note that vendors are asked an additional question regarding other ways that their recommendations regarding the measure might have been influenced. Their responses are not used in the direct calculation of the NTGR but are potentially useful in making adjustments to the core NTGR.

The Timing and Selection Score is calculated as:

The highest of the responses to the first four decision maker questions and, if the vendor interview has been triggered, the VMAX score multiplied by the score the decision makers assigned to the vendor recommendation.

5.1.2. Program Influence Score

The questions asked are:

1. Did you learn about PROGRAM BEFORE or AFTER you decided to implement the specific MEASURE that was eventually adopted or installed?
2. Now I'd like to ask you a last question about the importance of the program to your decision as opposed to other factors that may have influenced your decision. Again using the 0 to 10 rating scale we used earlier, where 0 means “Not at all important” and 10 means “Very important,” please rate the overall importance of PROGRAM versus the most important of the other factors we just discussed in your decision to implement the specific MEASURE that was adopted or installed. This time I would like to ask you to have the two importance ratings -- the program importance and the non-program importance -- total 10.

The Program Influence score is calculated as:

The importance of the program, on the 0 to 10 scale, to question 2. This score is reduced by half if the respondent learned about the program after the decision had been made.

5.1.3. No-Program Score

The questions asked are:

1. Regarding the installation of this equipment, if the PROGRAM had not been available, using a likelihood scale from 0 to 10, where 0 is “Not at all likely” and 10 is “Extremely likely” how likely is it that you would have installed exactly the same item/equipment, using a 0 to 10 scale, where 0 is not at all likely and 10 is extremely likely?

2. IF 1>0. You indicated that there was an “X” in 10 likelihood that you would have installed the same equipment if the PROGRAM had not been available. When do you think you would have installed this equipment? Please express your answer in months
 - a. _____ within 6 months? (Deferred NTG Value=0)
 - b. _____ 7 to 47 months later (Deferred NTG Value=(months-6)*.024)
 - c. _____ 48 or more months later (Deferred NTG Value =1)
 - d. _____ Never (Deferred NTG Value=1)

Note: The value 0.024 is 1 divided by 41 (41 is calculated as 47 – 6). This assumes that the deferred NTG value is a linear function beginning in month 7 through month 47, increasing 0.024 for each month of deferred installation.

The No-Program Score is calculated as:

10 minus (the likelihood of installing the same equipment multiplied by one minus the *deferred net-to-gross value* associated with the timing of that installation).

5.1.4. The Core NTGR

The self-reported core NTGR in most cases is simply the average of the Program Influence, Timing and Selection, and No-Program Scores, divided by 10. The one exception to this is when the respondent indicates a 10 in 10 probability of installing the same equipment at the same time in the absence of the program, in which case the NTGR is based on the average of the Program Influence and No-Program scores only.

5.2. Data Analysis and Integration

The calculation of the Core NTGR is fairly mechanical and is based on the answers to the closed-ended questions. However, the reliance of the Standard NTGR – Very Large on more information from so many different sources requires more of a case study level of effort. The SRA Guidelines point out that a case study is one method of assessing both quantitative and qualitative data in estimating a NTGR. A case study is an organized presentation of all these data available about a particular customer site with respect to all relevant aspects of the decision to install the efficient equipment. In such cases where multiple interviews are conducted eliciting both quantitative and qualitative data and a variety of program documentation has been collected, one will need to integrate all of this information into an internally consistent and coherent story that supports a specific NTGR.

The following data sources should be investigated and reviewed as appropriate to supplement the information collected through the decision maker interviews.

- Account Representative Interview
- Utility Program Manager/Staff Interview
- Utility Technical Contractor Interview
- Third party Program Manager Interview
- Evaluation Engineer Interview
- Gross Impact Site Plan/Analysis Review
- Corporate Green/Environmental Policy Review (if mentioned as important)
- Corporate Standard Practice Review (if mentioned as important)
- Industry Standard Practice Review (if mentioned as important)
- Corporate payback review (if mentioned as important)
- Review relevant codes and standards, including regulatory requirements
- Review industry publications, websites, reports such as the Commercial Energy Use Survey, historical purchase data of specific measures etc.

As detailed in the Self-Report NTGR Guidelines, when complementing the quantitative analysis of free-ridership with additional quantitative and qualitative data from multiple respondents and other sources, there are some basic concerns that one must keep in mind. Some of the other data – including interviews with third parties who were involved in the decision to install the energy efficient equipment – may reveal important influences on the customer’s decision to install the qualifying program measure. When one chooses to incorporate other data, one should keep the following principles in mind: 1) the method chosen should be balanced. That is, the method should allow for the possibility that the other influence can either increase or decrease the NTGR calculated from the decision maker survey responses, 2) the rules for deciding which customers will be examined for potential other influences should be balanced. In the case of Standard –Very Large interviews, all customers are subject to such a review, so that the pool of customers selected for such examination will not be biased towards ones for whom the evaluator believes the external influence will have the effect of influencing the NTGR in only one direction, 3) the plan for capturing other influences should be based on a well-conceived causal framework. The onus is on the evaluator to build a compelling case using a variety of quantitative and/or qualitative data for estimating a customer’s NTGR.

Establishing Rules for Data Integration

Before the analysis begins, the evaluation team should establish, to the extent feasible, rules for the integration of the quantitative and qualitative data. These rules should be as specific as possible and be strictly adhered to throughout the analysis. Such rules might include instructions regarding when the NTGR based on the quantitative data should be overridden based on qualitative data, how much qualitative data are needed to override the NTGR based on quantitative data, how to handle contradictory information provided by more than one person at a given site, how to handle situations when there is no

decision-maker interview, when there is no appropriate decision-maker interview, or when there is critical missing data on the questionnaire, and how to incorporate qualitative information on deferred free-ridership.

One must recognize that it is difficult to anticipate all the situations that one may encounter during the analysis. As a result, one may refine existing rules or even develop new ones during the initial phase of the analysis. One must also recognize that it is difficult to develop algorithms that effectively integrate the quantitative and qualitative data. It is therefore necessary to use judgment in deciding how much weight to give to the quantitative versus qualitative data and how to integrate the two. The methodology and estimates, however, must contain methods to support the validity of the integration methods through preponderance of evidence or other rules/procedures as discussed above.

For the **Standard-Very Large** cases in the large Nonresidential programs, the quantitative data used in the NTGR Calculator (which calculates the “core” NTGR), together with other information collected from the decision maker regarding the installation decision, form the initial basis for the NTG “story” for each site. Note that in most cases, supplemental data such as tracking data, program application files and results of interviews with program/IOU staff and vendors, will have been completed before the decision maker is contacted and will help guide the non-quantitative questioning in the interview. In practice, this means that most potential inconsistencies between decision maker responses and other sources of information should have been resolved before the interview is complete and data are entered into the NTGR Calculator. For example, if a company has an aggressive “green” policy widely promoted on its website that is not mentioned by the decision makers, the interviewer will ask the respondent to clarify the role of that policy in the decision. Conversely, if the decision maker attributes the decision to install the equipment to a new company wide initiative rather than the program, yet there is no evidence of such an initiative reported by program staff, vendors, or the company’s website, the decision maker will be asked to explain the discrepancy so that his or her responses can be changed if needed.

In some cases, however, it may be necessary to modify or override one of the scores contributing to the overall NTGR or the NTGR itself. Before this is done all quantitative and qualitative data will be systematically (and independently) analyzed by two experienced researchers who are familiar with the program, the individual site and the social science theory that underlies the decision maker survey instrument. Each will determine whether the additional information justifies modifying the previously calculated NTGR score, and will present any recommended modifications and their rationale in a well-organized manner, along with specific references to the supporting data. Again, it is important to note that the other influences can have the effect of either increasing or decreasing the NTGR calculated from the decision maker survey responses, and one should be skeptical about a consistent pattern of “corrections” in one direction or another.

Sometimes, *all* the quantitative and qualitative data will clearly point in the same direction while, in others, the *preponderance* of the data will point in the same direction. Other cases will be more ambiguous. In all cases, in order to maximize reliability, it is

essential that more than one person be involved in analyzing the data. Each person must analyze the data separately and then compare and discuss the results. Important insights can emerge from the different ways in which two analysts look at the same set of data. Ultimately, differences must be resolved and a case made for a particular NTGR. Careful training of analysts in the systematic use of rules is essential to insure inter-rater reliability⁴.

Once the individual analysts have completed their review, they meet to discuss their respective findings and present to the other the rationale for their recommended changes to the Calculator-derived NTGR. Key points of these arguments will be written down in summary form (e.g., Analyst 1 reviewed recent AQMD ruling and concluded that customer would have had to install the same measure within 2 years, not 3, thereby reducing NP score from 7.8 to 5.5) and also presented in greater detail in a workpaper so that an independent reviewer can understand and judge the data and the logic underlying each NTGR estimate. Equally important, the CPUC will have all the essential data to enable them to replicate the results, and if necessary, to derive their own estimates.

The outcome of the reconciliation by two analysts determines the final NTGR for a specific project. Again, the reasoning behind the “negotiated” final value must be thoroughly documented in a workpaper, while a more concise summary description of the rationale can be included in the NTGR Calculator workbook (e.g., Analyst 1 and Analyst 2 agreed that the NTGR score should have been higher than the calculated value of 0.45 because of extensive interaction between program technical staff and the customer, but they disagreed on whether this meant the NTGR should be .6 or .7. After discussion, they agreed on a NTGR of .65 as reflecting the extent of program influence on the decision).

In summary, it has been decided that supplemental data from non-core NTG questions collected through these surveys should be used in the following ways in the California Large Nonresidential evaluations:

- Vendor interview data will be used at times in the direct calculation of the NTGR. It will also be used to provide context and confirming/contradictory information for Standard-Very Large decision maker interviews.
- Qualitative and quantitative information from other sources (e.g., industry data, vendor estimates of sales in no-program areas, and other data as described above) may be used to alter core inputs only if contradictions are found with the core survey responses. Since judgments will have to be made in deciding which information is more compelling when there are contradictions, supplemental data are reviewed independently by two senior analysts, who then summarize their findings and recommendations and together reach a final NTGR value.

⁴ Inter-rater reliability is the extent to which two or more individuals (coders or raters) agree. Inter-rater reliability addresses the consistency of the implementation of a rating system.

- Responses will also be used to construct a NTGR “story” around the project; that is they will help to provide the context and rationale for the project. This is particularly valuable in helping to provide guidance to program design for future years. It may be, for example, that responses to the core questions yield a high NTGR for a project, but additional information sources strongly suggest that the program qualifying technology has since become standard practice for the firm or industry, so that free ridership rates in future years are likely to be higher if program rules are not changed.
- Findings from other non-core NTGR questions (e.g., Payback Battery, Corporate Policy Battery) are also be used to **cross-check the consistency** of responses to core NTGR questions. When an inconsistency is found, it is presented to the Decision Maker respondent who is then be asked to explain and resolve it if they can. If they are not able to do so, their responses to the core NTGR question with the inconsistency may be overridden by the findings from these supplemental probes. These situations are handled on a case-by-case basis; however consistency checks are programmed into the CATI survey instrument used for the Basic and Standard cases.

Finally, some analysis of additional information beyond the close-ended questions that are used to calculate the Core NTGR could be done for the **Standard NTGR**. For example information regarding the financial criteria used to make capital investments, corporate policy regarding the purchase of energy efficiency equipment or the influence of standard practice in the same industry as the participant could be taken into account and used to make adjustments to the Core NTGR in a manner similar what is done for the Standard – Very Large NTGR.

5.3. Accounting for Partial Free Ridership

Partial free-ridership can occur when, in the absence of the program, the participant would have installed something more efficient than the program-assumed baseline efficiency but not as efficient as the item actually installed as a result of the program.

In situations where there is partial free ridership, the assumed baseline condition is affected. Absent partial free ridership, the assumed baseline would normally be based on existing equipment (in early replacement cases), on code requirements (in normal replace on burnout cases), or on a level above current code (e.g., this could be a market average or value purposefully set above code minimum but below market average; in this case, the definition and requirement would typically be defined by a specific program’s baseline rules). In some cases, there may be a “dual” baseline (more specifically, a baseline that changes over the measure’s EUL) if the project involves early replacement plus partial free ridership. In such cases, the baseline basis for estimating savings is the existing equipment over the remaining useful life (RUL) of the equipment, and then a baseline of likely intermediate efficiency equipment (e.g., code or above) for the remainder of the analysis period (i.e., the period equal to the EUL-RUL). When there is partial free ridership, the baseline equipment that would have been installed absent the program is of an intermediate efficiency level (resulting in lower energy savings than that assumed by the program if the program took in situ equipment efficiency as the basis for

savings over the entire EUL). A related issue with respect to determination of the appropriate baseline is whether the adjustment made, if any, from the in situ or otherwise claimed baseline in the ex ante calculation, is whether the adjustment applies to the gross or net savings calculation.

Assignment of Partial Free Ridership Effects to Gross versus Net. In past evaluations, partial free ridership impacts have principally been incorporated into the net-to-gross ratio. This is because most partial free ridership is induced by market conditions, rather than by non-market factors. Market conditions refer primarily to standard adoption of a technology by a particular market segment or end user as a result of competitive market forces or other end user-specific factors. The key determining principle with respect to application of the adjustment to the net-to-gross ratio is whether there is a level of efficiency, below the efficiency of the measure for which savings are paid and claimed, but above what is required by code or minimum program baseline requirements that the end user would have implemented anyway without the program. Conditions that cause this adjustment to be made to gross savings rather than the net-to-gross ratio may include factors such as

- changing baseline equipment to meet changed business circumstances (such as increased production/throughput, changes in occupancy, etc.);
- compliance with environmental regulations, indoor air quality requirements, safety requirements; or
- the need to address an operational problem.

Each project should be examined separately for partial free ridership and a determination should be made based on the unique circumstances of each installation of whether an adjustment to gross savings or the net-to-gross ratio is warranted.

Data Collection Procedures. Information is gathered on partial free ridership using the following questions asked as part of the decision maker NTGR survey.

1. Now I would like you to think one last time about what action you would have taken if the program had not been available. Supposing that you had not installed the program qualifying equipment, which of the following alternatives would you have been MOST likely to do?
 - a. Install fewer units
 - b. Install standard efficiency equipment or whatever required by code
 - c. Install equipment more efficient than code but less efficient than what you installed through the program
 - d. repair/rewind or overhaul the existing equipment
 - e. do nothing (keep the existing equipment as is)
 - f. something else (specify what _____)
2. (IF FEWER UNITS) How many fewer units would you have installed? (It is okay to take an answer such as ...HALF...or 10 percent fewer ... etc.)

3. (IF MORE EFFICIENT THAN CODE) Can you tell me what model or efficiency level you were considering as an alternative? (It is okay to take an answer such as ... 10 percent more efficient than code or 10 percent less efficient than the program equipment)
4. (IF REPAIR/REWIND/OVERHAUL) How long do you think the repaired/rewound/refurbished equipment would have lasted before requiring replacement?

In addition, these same partial free ridership questions should be asked during the on-site audit for a given project. This latter interview will be conducted by the project engineers. The collected information helps the gross impact and NTG analysis teams gain a more complete understanding of the true project baseline and equipment selection decision. These decision maker questions are included in the Excel version of the CATI-based Standard and Basic decision maker survey instrument as well as in the Standard-Very Large instrument.

Data Analysis and Integration Procedures. In cases where partial free ridership is found and it is determined that the adjustment should be made to the net-to-gross ratio, the following procedure should be used:

On the net side, the adjustment is based on the intermediate baseline indicated by the decision maker for the time period in which the intermediate equipment would have been installed. The calculation of energy saved under this intermediate baseline is done, and then divided by the savings calculated under the in situ baseline. The resulting ratio is then multiplied by the initial NTGR which was previously calculated using only the 'core' scoring inputs. The effect of this adjustment is to reduce the NTGR further to reflect the effects of the revealed partial free ridership.

In all cases, the Gross Impacts and NTG analysis teams will need to carefully coordinate their calculations to ensure that they are not inadvertently adjusting the savings twice for the same partial free ridership, i.e., through adjustments both to the gross savings calculation and to the NTG ratio.

6. NTGR INTERVIEW PROCESS

The NTGR surveys are conducted via telephone interviews. Highly-trained professionals with experience levels that are commensurate with the interview requirements should perform these interviews. Basic and Standard level interviews should be conducted by senior interviewers, who are highly experienced conducting telephone interviews of this type. Standard - Very Large interviews should be completed by professional consulting staff due to the complex nature of these projects and related decision making processes. More than likely, these will involve interviews of several entities involved in the project including the primary decision maker, vendor representatives, utility account executives, program staff and other decision influencers, as well as a review of market data to help establish an appropriate baseline.

All but the Standard -Very Large interviews should be conducted using computer-aided telephone interview (CATI) software. Use of a CATI approach has several advantages: (1) the surveys can be customized to reflect the unique characteristics of each program, and associated program descriptions, response categories, and skip patterns; (2) it drastically reduces inaccuracies associated with the more traditional paper and pencil method; and (3) the process of checking for inconsistent answers can be automated, with follow up prompts triggered when inconsistencies are found.

7. COMPLIANCE WITH SELF-REPORT GUIDELINES

The proposed NTGR framework fully complies with all of the CPUC/ED and the MECT's Guidelines for Estimating Net-to-Gross Ratios Using the Self-Report Approach, as demonstrated in Appendix D.

Appendix B

References

- Blalock, H. (1970). Estimating measurement error using multiple indicators and several points in time," *American Sociological Review*, 35, pp. 101-111.
- Bogdan, Robert and Steven J. Taylor. (1975). *Introduction to qualitative research methods*. New York: John Wiley & Sons.
- Britan, G. M. (1978). Experimental and contextual models of program evaluation. *Evaluation and Program Planning*, 1: 229-234.
- Cochran, William G. (1977). *Sampling techniques*. New York: John Wiley & Sons.
- Crocker, L. and J. Algina. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.
- Cronbach L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- DeVellis, R.F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage Publications, Inc.
- Duncan, O.D. (1984). *Notes on social measurement: Historical and critical*. New York: Russell Sage.
- Guba, E. G. (1978). Toward a methodology of naturalistic inquiry in educational evaluation. *CSE Monographic Series in Evaluation No. 8*. Los Angeles: Center for the Study of Evaluation.
- Hall, Nick, Johna Roth, Carmen Best, Sharyn Barata, Pete Jacobs, Ken Keating, Ph.D., Steve Kromer, Lori Megdal, Ph.D., Jane Peters, Ph.D., Richard Ridge, Ph.D., Francis Trottier, and Ed Vine, Ph.D. (2007). *California Energy Efficiency Evaluation: Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*. Prepared for the California Public Utilities Commission.
- Lyberg, Lars, Paul Biemer, Martin Collins, Edith De Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin. (1997). *Survey measurement and process quality*. New York, NY: John Wiley & Sons.
- Madow, William G., Harold Nisselson, Ingram Olkin. (1983). *Incomplete data in sample surveys*. New York: Academic Press.
- Maxwell, Joseph A. (2004). Using Qualitative Methods for Causal Explanations. *Field Methods*, Vol. 16, No. 3, 243-264.

- Megdal, Lori, Yogesh Patil, Cherie Gregoire, Jennifer Meissner, and Kathryn Parlin (2009). Feasting at the Ultimate Enhanced Free-Ridership Salad Bar. *Proceedings of the International Energy Program Evaluation Conference*.
- Mohr, Lawrence B. (1995). *Impact analysis for program evaluation*. Thousand Oaks, CA: Sage Publications, Inc.
- Netemeyer, Richard G., William O. Bearden, and Subhash Sharma. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: SAGE Publications.
- Patton, Michael Quinn. (1987). *How to use qualitative methods in evaluation*. Newbury Park, California: SAGE Publications.
- Ridge, Richard, Philippus Willems, and Jennifer Fagan. (2009). Self-Report Methods for Estimating Net-to-Gross Ratios in California: Honest! *Proceedings from the 19th National Energy Services Conference*.
- Ridge, Richard, Philippus Willems, Jennifer Fagan and Katherine Randazzo. (2009). The Origins of the Misunderstood and Occasionally Maligned Self-Report Approach to Estimating the Net-To-Gross Ratio. *Proceedings of the International Energy Program Evaluation Conference*.
- Rogers, Patricia J., Timothy A. Hacsí, Anthony Petrosino, and Tracy A. Huebner (Eds.) (2000). *Program theory in evaluation: Challenges and opportunities*. San Francisco, CA: Jossey-Bass Publishers.
- Rossi, Peter and Howard E. Freeman. (1989). *Evaluation: A systematic approach*. Newbury Park, California: SAGE Publications.
- Sayer, Andrew. (1992). *Method in social science: A Realist Approach*. New York: Routledge.
- Sax, Gilbert. (1974). *Principles of educational measurement and evaluation*. Belmont, CA: Wadsworth Publishing Company, Inc.
- Schumacker, Randall E. and Richard G. Lomax. (1996). *A beginner's guide to structural equation modeling*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Scriven, Michael. (1976). Maximizing the power of causal explanations: The modus operandi method. In G.V. Glass (Ed.), *Evaluation Studies Review Annual, Vol. 1*, pp.101-118). Beverly Hills, CA: Sage Publications.
- Shadish, Jr., William R. and Thomas D. Cook, and Laura C. Leviton. (1991). *Foundations of program evaluation*. Newbury Park, CA: Sage Publications, Inc.
- Stone, Arthur A., Jaylan S. Turkkan, Christine A. Bachrach, Jared B. Jobe, Howard S. Kurtzman, and Virginia S. Cain. (2000). *The science of the self-report: Implications for research and practice*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Tashakkori, Abbas and Charles Teddlie. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: SAGE Publications.

TecMarket Works, Megdal & Associates, Architectural Energy Corporation, RLW Analytics, Resource Insight, B & B Resources, Ken Keating and Associates, Ed Vine and Associates, American Council for an Energy Efficient Economy, Ralph Prahl and Associates, and Innovologie. (2004). *The California evaluation framework*. Prepared for the California Public Utilities Commission and the Project Advisory Group.

Velleman, P. F., and Wilkinson, L. (1993), Nominal, ordinal, interval and ratio typologies are misleading. *American Statistician*, 47(1), 65-72.

Weiss, Carol H. (1998). *Evaluation*. Upper Saddle River, New Jersey: Prentice Hall.

Weiss, R. S. and M.Rein. (1972). The Evaluation of broad-aim programs: Difficulties in experimental design and an alternative. In C. H. Weiss (ed.) *Evaluating action programs: Readings in social action and education*. Boston: Allyn and Bacon.

Wholey, Joseph S., Harry P. Hatry and Kathryn E. Newcomer. (1994). *Handbook of practical program evaluation*. San Francisco, CA: Jossey-Bass, Inc.

Yin, Robert K. (1994). *Case study research: Design and methods*. Newbury Park, California: SAGE Publications.

Appendix B

Net-to-Gross Questions and Uses of Data by Level of NTGR Analysis

Note: A more detailed version of this survey, with skip patterns and complete response categories, is available in Excel format from the NTG Working Group or at <http://www.energydataweb.com/cpuc/default.aspx>

DECISION MAKER SURVEY

	Question Text	Basic	Standard and Standard – Very Large
	<p>Introduction</p> <p>Hello, my name is _____ from COMPANY NAME and I am calling about your recent participation in PROGRAM NAME. Are you the person who was most involved with the decision to participate in the PROGRAM NAME? [IF YES, CONTINUE]. We are interviewing firms that participated in the PROGRAM NAME in 2006 and 2007 to discuss the factors that may have influenced your decision to participate in the program. The interview will take about 20 minutes. The questions on this survey pertain to work completed by your company at this current address, excluding other locations.</p>		
	<p>WARM-UP QUESTIONS</p>		
A1	<p>First, according to our records, you participated in PROGRAM NAME on (approximate date). [READ: Program Description. PROGRAM NAME promotes energy efficiency improvements in commercial/industrial facilities. The program offers (choose all that apply): energy audits to help identify applicable measures, feasibility studies to analyze the energy and cost savings of recommended measures, incentives to help cover a portion of the cost of implementing energy efficient measures, etc. Is that correct?</p>	X	X
	Yes, No, DK, Refused		
A2	<p>Next, I'd like to confirm the following information regarding the measures you implemented through the program: (READ: PROJECT DETAILS INCLUDING SERVICES RECEIVED, MEASURES INSTALLED, KEY DATES, PARTICIPATING VENDORS, ETC.) Does that sound right?</p>	X	X
	Yes, No, DK, Refused		
A3	<p>Why did you decide to implement MEASURE NAME? Were there any other reasons?</p>	X	X
	a. Record VERBATIM		
	b. DK/Refused		
	<p>NET-TO-GROSS BATTERY</p>		
N1	<p>When did you first learn about PROGRAM? Was it BEFORE or AFTER you first began to think about implementing MEASURE?</p>	X	X
	a. Before (Skip to N3)		
	b. After		
	c. DK/Refused		

N2	Did you learn about PROGRAM BEFORE or AFTER you decided to implement the specific MEASURE that was eventually adopted or installed?	X	X
	a. Before		
	b. After		
	c. DK/Refused		
	<i>READ: Program Description: As I mentioned earlier, [PROGRAM NAME] promotes energy efficiency improvements in commercial/industrial facilities. The program offers (choose all that apply): energy audits to help identify applicable measures, feasibility studies to analyze the energy and cost savings of recommended measures, incentives to help cover a portion of the cost of implementing energy efficient measures, etc. I'm going to ask you to rate the importance of the program as well as other factors that might influence your decision to implement [MEASURE.] Think of the degree of importance as being shown on a scale with equally spaced units from 0 to 10, where 0 means not at all important and 10 means very important, so that an importance rating of 8 shows twice as much influence as a rating of 4.</i>		
N3	Now, using this 0 to 10 rating scale, where 0 means "Not at all important" and 10 means "Very important," please rate the importance of each of the following in your decision to implement this specific [MEASURE] at this time. [CUSTOMIZE LIST OF FACTORS FOR PROGRAM BEFORE ASKING THEM TO SCORE THE FULL LIST. ROTATE PRESENTATION OF ITEMS. FOLLOW UP WITH "And is there anything else that I may have missed?" RECORD AS p. Other (SPECIFY)]		
	a. The age or condition of the old equipment	X	X
	b. Availability of the PROGRAM rebate	X	X
	c. Information provided through a recent feasibility study, energy audit or other types of technical assistance provided through the PROGRAM (probe on when and by whom?)	X	X
	d. Recommendation from a vendor/supplier (If >5, Vendor interview may be triggered)	TRIGGER	TRIGGER
	e. Previous experience with PROGRAM?	X	X
	f. Previous experience with this MEASURE?	X	X
	g. Information from PROGRAM training course?	X	X
	h. Information from other PROGRAM marketing materials?	X	X
	i. A recommendation from an auditor or consulting engineer	X	X
	j. Standard practice in our business/industry (IF >5, ask standard practice battery)	X	TRIGGER
	k. Endorsement or recommendation by PROGRAM staff, PROGRAM vendor, or UTILITY representative	X	X
	l. Corporate policy or guidelines (If >5 ask Policy questions)	X	TRIGGER
	m. Payback on the investment (If >5 ask payback battery)	X	TRIGGER
	n. General concerns about the environment	X	X
	o. Specific concerns about global warming	X	X
	p. Specific concerns about achieving energy independence	X	X
	q. Other (SPECIFY)	X	X
N4	Now I'd like to ask you a last question about the importance of the program to your decision. Again using the 0 to 10 rating scale we used earlier, where 0 means "Not at all important" and 10 means "Very important," please rate	X	X

	the overall importance of PROGRAM versus the other factors we just discussed in your decision to implement the specific MEASURE. I'd like you to give me a 0 to 10 score for the PROGRAM's influence and a 0 to 10 score for the influence of the most important other factor so that the two scores total 10.		
	a. _____ rating of the importance of PROGRAM NAME	X	X
	b. _____ rating of the importance of Other Factors	X	X
	<i>Now I would like you to think about the action you would have taken with regard to the installation of this equipment PROGRAM had not been available.</i>		
N5	Regarding the installation of this equipment if the PROGRAM had not been available, how likely is it that you would have installed exactly the same item/equipment, using a 0 to 10 likelihood scale, where 0 is not at all likely and 10 is extremely likely?	X	X
N6	<i>IF N5>0.</i> You indicated in your previous responses that there was a X in 10 likelihood that you would have installed the same equipment if the PROGRAM had not been available.	X	X
	When do you think you would have installed this equipment? (Please answer in months) _____		
	a. _____ ..within 6 months? NTGR = 0		
	b. _____ .. 6 – 47 months later (NTGR=(months-6)*.024)		
	c. _____ ..4 or more years later (NTGR=1)		
	g. _____ ..Never (NTGR=1)		
	PARTIAL FREE RIDERSHIP BATTERY	GROSS IMPACT	GROSS IMPACT
P1	Now I would like you to think one last time about what action you would have taken if the program had not been available. Supposing that you had not installed the program qualifying equipment, which of the following alternatives would you have been MOST likely to do?: a. Install fewer high efficiency units (e.g., controls, VFDs, lights) b. Install standard efficiency equipment or whatever required by code c. Install equipment more efficient than code, but less efficient than we installed through the program d. Repair/rewind/refurbish the existing equipment e. do nothing (keep the existing equipment as is) f. Something else (specify)		
P4	If P1=a: How many units would you have installed? Record number of units or percentage of units actually installed		
P5			
P6	If P1=c: Can you tell me what model or efficiency level you were considering as an alternative? (It is okay to take an answer such as ... 10 percent more efficient than code or 10 percent less efficient than the program equipment)		
P7	If P1=d: How long do you think the repaired/rewound/refurbished equipment would have lasted before requiring replacement?		
P8			
P9			
	Additional Decision Maker Questions		

	PAYBACK BATTERY (If payback importance >5)		
N10	What financial calculations does your company make before proceeding with installation of a MEASURE like this one?		X
N11	What is the cut-off point your company uses before deciding to proceed with the investment?		X
N12	What was the result of the calculation for MEASURE: a) with the rebate? b) without the rebate?		X
	<i>INVESTIGATE INCONSISTENT RESPONSE</i>		
N13	What competing investments, if any, were considered for the funds that were allocated to the adoption of MEASURE?		X
N14	Why was MEASURE chosen over these other investments		X
	CORPORATE POLICY BATTERY (If corporate policy importance >5)		
N15	Does your organization have a corporate environmental policy to reduce environmental emissions or energy use? Some examples would be to "buy green" or use sustainable approaches to business investments.		X
N16	What specific corporate policy influenced your decision to adopt or install MEASURE?		X
N17	Had that policy caused you to adopt the MEASURE at this facility before participating in this program?		X
N18	Had that policy caused you to adopt the MEASURE at other facilities before participating in this program? When and where?		X
N19	Did you receive an incentive for a previous [MEASURE]? If so, please describe.		X
	STANDARD PRACTICE BATTERY (If standard practice importance >5)		
N20	How long has MEASURE been standard practice in your industry?		X
N21	Does your company ever deviate from the standard practice? If yes, under what conditions?		X
N22	How did this standard practice influence your decision to install the energy efficiency equipment		X
N23	What industry group or trade organization do you look to establish standard practice for your industry?		X
N24	How do you and other firms/facilities receive information on updates in standard practice?		X
	OTHER INFLUENCES BATTERY		
N25	Who provided the most assistance in the design or specification of MEASURE? Designer or Consultant, Equipment Distributor or Mfr Rep, Installer, Utility rep, or Internal staff	X	X
N26	Please describe the type of assistance that they provided.	X	X
N27	Please state, in your own words, any other factors that influenced your decision to go ahead on this energy efficient equipment/project.	X	X

VENDOR SURVEY

	Question Text	Basic	Standard and Standard Very Large
	Warm Up		
A1	The CUSTOMER indicates that you recommended the installation of [EFFICIENT MEASURE] at their facility at [CUSTOMER LOCATION] on [DATE]. Do you recall making this recommendation?	X	X
	a. Yes		
	b. No		
	c. DK (-8)		
	d. Refused (-9)		
	<i>I'm going to ask you to rate the importance of the [PROGRAM] in influencing your decision to recommend [MEASURE] to [CUSTOMER] and other customers. Think of the degree of importance as being shown on a scale with equally spaced units from 0 to 10, where 0 means not at all important and 10 means very important, so that an importance rating of 8 shows twice as much influence as a rating of 4.</i>		
V1	Using this 0 to 10 scale where 0 is "Not at all important" and 10 is "Very Important", how important was PROGRAM, including incentives as well as program services and information, in influencing your decision to recommend that CUSTOMER install the energy efficiency MEASURE at this time?	X	X
V2	And using a 0 to 10 likelihood scale, where 0 denotes "not at all likely" and 10 denotes "very likely," if the PROGRAM, including incentives as well as program services and information, had not been available, what is the likelihood that you would have recommended this specific energy efficiency MEASURE to CUSTOMER?	X	X
V3	Now, using a 0 to 100 percent scale, in what percent of sales situations did you recommend MEASURE before you learned about the [PROGRAM]?	X	X
V4	And using the same 0 to 100 percent scale, in what percent of sales situations do you recommend MEASURE now that you have worked with the [PROGRAM]?	X	X
V4a	In what other ways have your recommendations regarding MEASURE been influenced? [For each mention, ask: And using the same 0 to 10 scale, where 0 is "Not at all important" and 10 is "Very important", how important in influencing your recommendations. . . (INSERT FIRST MENTION, INSERT SECOND MENTION ETC.)]	X	X
V5	And, using the same 0 to 10 scale where 0 is "Not at all important" and 10 is "Very important", how important in your recommendation were		
	a. Training seminars provided by UTILITY?	X	X
	b. Information provided by the UTILITY website?	X	X
	c. Your firm's past participation in a rebate or audit program sponsored by UTILITY?	X	X

	Optional:		
V6	Approximately what percentage of your sales of MEASURE in UTILITY'S service territory are energy efficient models that qualify for incentives from the UTILITY program.	X	X
V7	On a 0 percent to 100 percent scale, in what percent of sales situations do you encourage your customers in UTILITY territory to purchase program qualifying [MEASURES]?	X	X
V8.	(IF LESS THAN 100) In what situations do you NOT encourage your customers to purchase energy efficient models if they qualify for a rebate? Why is that?	X	X
V9	Of those installations of EQUIPMENT in UTILITY service territory that qualify for incentives, approximately what percentage do not receive the incentive?	X	X
V10	Why do they not receive the incentive (open end?)	X	X
V11	Do you also sell MEASURE in areas where customers do not have access to incentives for energy efficient models?	X	X
V12	About what percent of your sales of MEASURE are represented by these areas where incentives are not available?	X	X
V12a	IF AT LEAST 10%: And approximately what percentage of your sales of MEASURE in these areas are the energy efficient models that would qualify for incentives in UTILITY'S service territory?	X	X
V13	Have you changed your stocking practices as a result of the UTILITY program? If yes, how?	X	X
V14	Do you promote energy efficient models equally in areas with and without incentives?	X	X

Appendix C

NTGR Scoring Algorithm and Example

The calculation of the self-report-based core NTGR is described below. The NTGR is calculated as an average of three scores representing responses to one or more questions about the decision to install a program measure.

1. A **Timing and Selection** score that captures the influence of the most important of various program and program-related elements in influencing the customer to select the specific program measure at this time. Program influence through vendor recommendations is also captured in this score.
2. An overall **Program Influence** score that captures the perceived importance of the program (whether rebate, recommendation, or other information) in the decision to implement the specific measure that that was eventually adopted or installed. The overall program influence score is reduced by half if the respondent says they learned about the program only after they decided to install the program qualifying measure.
3. A **No-Program** score that captures the likelihood of various actions the customer might have taken at this time and in the future if the program had not been available. This score accounts for deferred free ridership by capturing the likelihood that the customer would have installed program qualifying measures at a later date if the program had not been available.

Calculation of each of the above scores is discussed below. For each score, the questions contributing to the calculation are presented, the calculation is described, and an example is provided.

Timing and Selection Score

For the decision maker, the questions asked are:

Using a 0 to 10 rating scale, where 0 means not at all important and 10 means very important, please rate the importance of each of the following in your decision to implement this specific measure at this time:

- Availability of the PROGRAM rebate
- Information provided through a recent feasibility study, energy audit or other types of technical assistance provided through the PROGRAM
- Information from PROGRAM training course
- Information from other PROGRAM marketing materials
- Recommendation from a vendor/supplier (If >5, a vendor interview is triggered)

For the vendor, the questions asked if the interview is triggered are:

1. On a 0 to 10 scale where 0 is “Not at all important” and 10 is “Very important”, how important was PROGRAM, including incentives as well as program services and information, in influencing your decision to recommend that CUSTOMER install the energy efficiency MEASURE at this time?
2. And using a 0 to 10 likelihood scale, where 0 denotes “Not at all likely” and 10 denotes “Extremely Likely,” if the PROGRAM, including incentives as well as program services and information, had not been available, what is the likelihood that you would have recommended this specific energy efficiency MEASURE to CUSTOMER?
3. Now, using a 0 to 100 percent scale, in what percent of sales situations did you recommend this MEASURE before you learned about the PROGRAM?
4. And using the same 0 to 100 percent scale, in what percent of sales situations do you recommend this MEASURE now that you have worked with the PROGRAM?
5. And, using the same 0 to 10 scale where 0 is “Not at all important” and 10 is “Extremely Important”, how important in your recommendation were:
 - a. Training seminars provided by UTILITY?
 - b. Information provided by the UTILITY website?
 - c. Your firm’s past participation in a rebate or audit program sponsored by UTILITY?

If the vendor interview is triggered, a score is calculated that captures the highest degree of program influence on the vendor’s recommendation. This score (VMAX) is calculated as the MAXIMUM value of the following:

1. The response to question 1
2. 10 minus the response to question 2
3. The response to question 4 minus the response to question 3, divided by 10
4. The response to question 5 a.
5. The response to question 5b.
6. The response to question 5c.

The Timing and Selection Score is calculated as:

The highest of the responses to the first four decision maker questions and, if the vendor interview has been triggered, the VMAX score multiplied by the score the decision makers assigned to the vendor recommendation..

Example:

The decision maker provides responses of 5 for the importance of the rebate, 6 for an audit or feasibility study, 3 for training, 2 for other marketing materials, and 7 for the vendor recommendation, which means a vendor interview is triggered.

The vendor responses are 8 for the significance of the program, 5 for the likelihood of recommending the measure in the absence of the program, 40% for how often the measure was recommended before program awareness and 60% for how often it is recommended after program awareness, 3 for the importance of training, 2 for the importance of the website and 5

for the importance of previous participation. The VMAX score is the greatest of 8, (10-5), (60-40)/10, 3, 2 and 5. So VMAX is 8. This score is multiplied by the importance of the vendor recommendation, to which the decision maker assigned a 7, so the vendor score is 5.6.

The timing and selection score is the maximum of the four decision maker responses (5, 6, 3, and 2) and the vendor score (5.6). Even though the vendor interview was triggered, the vendor score is not as high as the 6 assigned to the importance of the audit or feasibility study, so the timing and selection score is 6.

Program Influence Score

The questions asked are:

1. Did you learn about PROGRAM BEFORE or AFTER you decided to implement the specific MEASURE that was eventually adopted or installed?
2. Again using the 0 to 10 rating scale we used earlier, where 0 means "Not at all important" and 10 means "Very important," please rate the overall importance of PROGRAM versus the most important of the other factors we just discussed in your decision to implement the specific MEASURE that was adopted or installed. This time I would like to ask you to have the two importance ratings -- the program importance and the non-program importance -- total 10.

The program influence score is calculated as:

The program importance response, on the 0 to 10 scale, to question 2. This score is reduced by half if the respondent became aware of the program only after having decided to adopt the program qualifying measure.

Example:

The decision maker says they became aware of the program before deciding to implement the measure, and provides a response of 7 to question 2, which becomes the program influence score.

No-Program Score

The questions asked are:

1. Regarding the installation of this equipment if the PROGRAM had not been available, how likely is it that you would have installed exactly the same item/equipment, using a 0 to 10 likelihood scale, where 0 is not at all likely and 10 is extremely likely?
2. IF 1>0. You indicated in your previous responses that there was an "X" in 10 likelihood that you would have installed the same equipment if the PROGRAM had not been available. When do you think you would have installed this equipment? Please express your answer in months
 - a. _____ Within 6 months? (Deferred NTG Value=0)
 - b. _____ 7 to 47 months later (Deferred NTG Value=(months-6)*.024)

- c. _____ 48 or more months later (Deferred NTG Value =1)
d. _____ Never (Deferred NTG Value=1)

Note: The value 0.024 is 1 divided by 41 (41 is calculated as 47 – 6). This assumes that the deferred NTG value is a linear function beginning in month 7 through month 47, increasing 0.024 for each month of deferred installation.

The No-Program Score is calculated as:

10 minus (the likelihood of installing the same equipment multiplied by one minus the deferred net-to-gross value associated with the timing of that installation).

Example

The respondent says there is a 4 in 10 likelihood that they would have installed the same equipment. In response to question 5, the decision maker says they would have installed the qualifying equipment 18 months later, which has a NTGR value of $(18-6)*.024$, or .29 associated with it.

The No-Program score is 10 minus $(4*(1-.29))$, which is 10 minus $4*.71$ or 7.16.

Core NTG Ratio

The self-reported core NTGR in most cases is simply the average of the Program Influence, Timing and Selection, and No-Program Scores, divided by 10. The one exception to this is when the respondent indicates a 10 in 10 probability of installing the same equipment at the same time in the absence of the program, in which case the NTGR is based on the average of the Program Influence and No-Program scores only.

Example (Core NTGR)

The NTGR is the average of 6, 8 and 7.2, or 7.1 divided by 10 = .71. This figure is then applied to adjusted gross savings to yield net savings.

Appendix D

Demonstration of Compliance with the CPUC/ED and MEC's Guidelines for Estimating Net-to-Gross Ratios Using the Self-Report Approach

1. Timing of the interview

To minimize problems of recall, every effort should be made to conduct the NTGR interview as close to project completion as possible.

2. Identifying the correct respondent

The survey form includes some initial probing on the respondent's role in the completed project, to confirm their involvement in the decision to implement the energy efficiency measures. In addition, both the utility or third party representative and any trade allies involved should be asked to confirm they are the correct contact. If multiple decision makers are identified, each one should be interviewed and the results pooled.

In the unfortunate circumstance where the key decision maker has left the company, that sample point should be discarded and replaced with a respondent from within the same stratum in the backup sample.

3. Set-up questions

The survey includes a series of warm-up questions that serve to remind the respondent about the circumstances and motivations surrounding the project, the project scope (including installed measures), incentives paid, and the project schedule. This information also helps to build the "story" to substantiate the NTGR responses given.

4. Use of multiple questions

The NTGR scoring algorithm relies on responses from several questions to determine the final NTGR score. The scoring is a function of:

- The timing of their program awareness relative to their decision to implement the installed measure
- The importance of program versus non-program influences in their decision making
- The importance of specific influences in the participant's general decision to implement the measure and that led them to implement the specific measure at the time they did rather than an alternative
- Without the program, the probability of alternative actions to implementing the selected measure

5. Validity and reliability

The proposed NTGR method is designed to produce valid and reliable NTGR results, based on the use of:

- *"Tried and true" question wording.* Many of the core questions used in NTGR scoring are substantially the same as those that have been used extensively in previous large C&I program evaluations, such as the last several rounds of evaluation for the California Standard Performance Contracting Program. While the question construct is somewhat

different from in the past, the wording used is essentially the same as has been used previously.

- *Information from supplemental questions and multiple data sources to corroborate and triangulate on the NTGR “story”.* In addition to self-reported information, the NTGR findings for Standard and Standard – Very Large NTGR sites include responses to a number of supplemental questions surrounding the project (e.g., corporate policy, standard industry practice and payback), and the results from an interview with the vendor(s) involved in the project. These findings will be used to converge on a plausible estimate of the NTGR and to help tell the “story” behind the project and its context.
- *Multiple reviewers. Standard - Very Large customer projects are reviewed by two experienced analysts.* The two reviewers seek to develop a NTGR consensus on the project, and resolve any differences of opinion.
- *Identification and explicit consideration of alternate hypotheses.* Respondents are asked about the relative influence of a variety of program and non-program factors.

During the pre-test of the NTGR survey instrument, reliability tests should be conducted using the CATI software. Any problem areas detected should be corrected.

6. Consistency checks

Questions within the NTGR battery that are more likely to produce inconsistent responses have been flagged. These include questions regarding the program’s reported importance in the decision to implement the specified measure, alternative actions in the program’s absence, questions reporting the motivations for doing the project, as well as any closely related supplemental questions. The CATI software should be specifically programmed to flag any inconsistencies, and include follow-up prompts when they are found. Interviewers should be instructed how to administer these follow-up questions to resolve these inconsistencies. Interviewers should make every effort to resolve any inconsistencies before concluding the interview. Examples of the procedures for checking consistency of responses are provided in Section 3.

7. Making the Questions Measure-Specific

In general, most projects involve one type or class of measure. However, there are a few instances where the project consists of multiple types of measures, but usually, one measure predominates. In such cases, the interview should be conducted around the dominant measure with the greatest share of savings. If there are projects with multiple types of measures and no one measure class predominates, the NTGR sequence should be repeated for each significant measure class (e.g., once for lighting and once for process measures). At the beginning of each interview, there is a prompt with a description of the measure class that the questions pertain to so that it is clear in the minds of the respondent which measures they are being asked about.

8. Partial free-ridership

Questions P1-P9 are designed to collect the information necessary to adjust for any partial free-ridership. *However, this adjustment is be made to the **gross savings** estimates and not to the NTGR.*

9. Deferred free-ridership

Question N6 addresses deferred free ridership, and provides specific adjustment factors for each response category. The NTGR algorithm (See Section 5 and Appendix C) text fully explains the specifics of this adjustment.

10. Scoring algorithms

The methodology includes a specific algorithm for developing a NTGR based on responses received. The results of the 0 to 10 scoring are used to develop specific values for each question used to score the NTGR. A description of the scoring algorithm is provided in Section 5 and in Appendix C.

11. Handling unit and item non-response

Every effort should be made to discourage non-responses (i.e., refusals and terminates). For example, in California, the interviewer points out that the energy efficiency program requires the project to be evaluated as a condition of participation. Absent such a requirement, interviewers should stress such things as the importance of evaluation in improving program design and delivery. In some cases, incentives can be offered to respondents. In the event various strategies are not successful, the non-responding customer should be replaced by another customer within the same stratum. While efforts to minimize item non-response (“don’t knows” and “refusals”) should be made using a variety of available techniques, one should recognize that forcing a response can distort the respondent’s answer and introduce bias.

12. Weighting the NTGR

The mean NTGR for a given measure, end use or program should be weighted to take into account the size of the ex post gross impacts.

13. Ruling out rival hypotheses

The core NTGR questions, particularly question 4 of the Decision Maker survey, have been carefully constructed to try to rule out rival hypotheses. The method asks respondents to jointly consider and rate the importance of the many likely events or factors that may have influenced their energy efficiency decision making, rather than focusing narrowly on only their rating of the program’s importance. This question structure more accurately reflects the complex nature of the real-world decision making and should help to ensure that all non-program influences are reflected in the NTGR assessment in addition to program influences.

14. Precision of the NTGR

The calculation of the achieved relative precision of the NTGRs (for program-related measures and practices and non-program measures and practices) is expected to be straightforward. However, the inclusion of more complicated situations involving multiple participant and vendor

interviews as well as the inclusion of additional qualitative information means that the NTGR standard errors may underestimate the uncertainty surrounding the NTGR estimate.

15. Pre-testing the questionnaire

The NTGR survey should be carefully and extensively pre-tested and adjusted in response to pre-test findings before it is fielded.

16. Incorporation of additional qualitative and quantitative data in estimating the NTGR (data collection, rules for data integration, analysis)

Specific rules have been established for data integration and these are described in Section 3.

17. Qualified interviewers

The NTGR surveys should be fielded by highly experienced interviewers. High level professional interviewers should be used for the largest and most complex projects, while less experienced professional interviewers should be used for smaller, simpler projects. A CATI approach should be used for all but the very largest and most complex projects.