TETRA TECH

complex world | CLEAR SOLUTIONS™

Massachusetts Program Administrators

Cross-Cutting C&I Free-Ridership and Spillover Methodology Study Final Report

April 18, 2011

KEMA

NMR Group, Inc.

Massachusetts Program Administrators

Cross-Cutting C&I Free-Ridership and Spillover Methodology Study Final Report

April 18, 2011

Revised May 20, 2011: per request of Massachusetts Program Administrators to add an Executive Summary

Prepared for: Massachusetts Program Administrators

Tetra Tech
6410 Enterprise Lane, Suite 300 | Madison, WI 53719
Tel 608.316.3700 | Fax 608.661.5181
www.tetratech.com

TABLE OF CONTENTS

1. Executive Summary 1-1
- 1.1 Study objectives 1-1
- 1.2 Background 1-1
- 1.3 Task 1: Kick-off meeting and program review 1-2
- 1.4 Task 2: Best practice review 1-2
- 1.5 Task 3: Develop and test methodology to measure free-ridership and spillover 1-4
- 1.6 Task 4: Presentation of the proposed final methodology and final report 1-4

2. Introduction 2-1
- 2.1 Introduction 2-1
- 2.2 Background 2-2
- 2.3 Study Objectives 2-3
- 2.4 Scope of the Research 2-3
  - 2.4.1 Task 1: Kick-off meeting and program review 2-3
  - 2.4.2 Task 2: Best practice review 2-3
  - 2.4.3 Task 3: Develop and test methodology to measure free-ridership and spillover 2-4
  - 2.4.4 Task 4: Presentation of the proposed final methodology and final report 2-4
- 2.5 Overview of Report 2-4

3. Methodologies for Studying Free-Ridership and Spillover 3-1
- 3.1 Types of Methodologies and their Advantages/Disadvantages 3-1
  - 3.1.1 Overview 3-1
  - 3.1.2 Discussion of general approaches 3-2
  - 3.1.3 Summary of methods 3-9
  - 3.1.4 Design and data collection considerations 3-11
- 3.2 Decision Framework for Selecting Appropriate Methodologies 3-12
  - 3.2.1 General 3-12
  - 3.2.2 Methods applicable for different conditions 3-12
  - 3.2.3 Application of decision framework to Massachusetts C&I programs 3-14

4. Literature Review on Self-Report Approach (SRA) Methodologies 4-1
- 4.1 Best Practices in Self-Report Approach (SRA) Methodologies 4-1
  - 4.1.1 Necessary data elements for implementation of SRA free-ridership and spillover measurement 4-1
  - 4.1.2 Elements of good design for self-report free-ridership and spillover measurement 4-2
  - 4.1.3 2003 standardized methodology 4-4
  - 4.1.4 Recommended changes to the 2003 standardized SRA methodology 4-5

5. Develop and Test Proposed Methodology 5-1
- 5.1 Free-Ridership Methodology Pretest 5-1
  - 5.1.1 Summary of the 2003 Standardized Methodology 5-2
  - 5.1.2 Summary of the 2010 Proposed Methodology 5-4
  - 5.1.3 2010 Methodology Pretest Survey Questions 5-5
  - 5.1.4 2010 Methodology Pretest Scoring 5-7
- 5.2 2010 Methodology Pretest Data Collection 5-11
  - 5.2.1 Sampling 5-11
  - 5.2.2 Results 5-12
  - 5.2.3 Different Scoring Scenarios 5-13
  - 5.2.4 Stability Analysis 5-14
- 5.3 Simulated Pretest Dataset Analysis 5-18
- 5.4 Spillover Pretest Methodology 5-20

5.4.1 "Like" Spillover 5-22
5.4.2 "Unlike" Spillover 5-22

6. Recommended Final Methodology 6-1

APPENDIX A: Taxonomy of PA Programs by Sector, Type of Assistance, Eligibility, Incentives, and Delivery A-1

APPENDIX B: Best Practice Review Sources of Information B-1

APPENDIX C: Bibliography of References C-1

APPENDIX D: Comparison of Median and Mean Substitutions D-1

APPENDIX E: Participant and Vendor Surveys for Full-scale Study E-1

APPENDIX F: Scoring Flowcharts F-1

1. EXECUTIVE SUMMARY

1.1 STUDY OBJECTIVES

In 2010 and 2011, the study team (Tetra Tech, KEMA, and NMR) conducted a review of the methodologies used to assess free-ridership and spillover for Commercial & Industrial (C&I) programs. The focus of this study was on the general methods for estimating what would have happened absent C&I programs in Massachusetts. The net program effect is the observed effect, less the estimate of what would have happened absent the program. The objectives of this study were to develop a standardized methodology for situations where C&I end-users are able to report on program impacts via self-report methods, and to provide a decision framework and guidelines for when the standardized self-report methodology is appropriate and when other methods need to be used (e.g., upstream programs).

1.2 BACKGROUND

Over the past decade, there has been extensive debate across the country regarding the need for and measurement of estimating what would have happened absent a program. Isolating the effects of these program factors, and other influences in the decision to adopt energy efficiency measures, is often referred to as "attribution." The increasing importance of program attribution given aggressive savings targets and DSM incentive mechanisms has intensified this debate. Some argue that free-ridership and spillover cancel each other out and should not be measured, that they are too difficult to reliably estimate, or that funds are better spent for program implementation. However, other evaluators and regulators note the advantages of consistent measurement of free-ridership and spillover. Fagan, Messenger, Rufo and Lai (2009)[1], list the following reasons why understanding a program's net savings is important:

- Understand program and portfolio cost-effectiveness
- Improve portfolio design and resource allocation
- Refine program design
- Understand market transformation
- Align program administrators' financial interests with societal interests
- Understand how energy efficiency programs affect baseline load forecasts and short-term power procurement decisions.

Prior to 2003, the Massachusetts Program Administrators (PAs) independently quantified the impacts of free-ridership and spillover (net-to-gross factors) for their energy efficiency programs targeting C&I customers in Massachusetts. These independent evaluation approaches resulted in reported impacts that were measured using a variety of survey instruments, analysis techniques, and assumptions.

In 2003, a consortium of Massachusetts PAs funded a study to develop standardized sampling techniques, data collection approaches, survey questions, survey instruments, and an analysis methodology that each of the PAs could use to determine free-ridership and spillover factors for C&I

[1] "A Meta-Analysis of Net to Gross Estimates in California". Jennifer Fagan, Mike Messenger, Mike Rufo, Peter Lai, paper presented at the 2009 AESP conference.

programs.[2] This standardization, which relied on customer and vendor self-reports, was important to the study sponsors for achieving consistency in the annual program results they report to state regulatory agencies.

For the current study, the team conducted four tasks to develop a revised standardized methodology, decision framework, and guidelines for C&I programs. These tasks included a kick-off meeting and program review, a best practice review, development and pretest of a revised methodology, and presentation of final recommendations for a revised methodology.

1.3 TASK 1: KICK-OFF MEETING AND PROGRAM REVIEW

As part of Task 1, the study team and key stakeholders from each PA discussed the objectives for the methodology study. The team then presented an overview of the existing 2003 methodology and its limitations (e.g., reliance on self-reports, generous handling of acceleration, conservative spillover measurement, limited market effects). Finally, we discussed potential issues for consideration in revising the methodology for C&I programs (e.g., changes in behavior, changes in O&M practices) and the tasks to be completed as part of this study.

As part of the program review, the study team reviewed the three-year plans and information already collected from the PAs by the KEMA team as part of the current C&I evaluations being conducted under separate contracts. Where additional information needed to be collected to develop a complete understanding of the programs (or specific measures within programs), and to develop the decision framework and methodology, the cross-cutting study team attempted to re-contact each PA to ask additional questions. The product of this review was a taxonomy of programs by sector, type of assistance, eligibility, incentives, and delivery. It was used to determine the type and number of approaches for review and development.

1.4 TASK 2: BEST PRACTICE REVIEW

This methodology study included a review of other methodologies being used across the nation. The review explored the pros and cons of alternative methods for estimating what would have happened in the absence of the program in different contexts. While the objective of this current study was to develop a standardized methodology for situations where end-users are able to report on program impacts via self-report methods, we recognize that some programs cannot be addressed using a standardized methodology. Therefore, as part of this task we developed a decision framework and guidelines for when the standardized self-report methodology is appropriate and when other methods need to be used (e.g., upstream programs).

As part of the best practice review, we reviewed and synthesized recent attribution research and compared it to the existing standardized methodology. We also reviewed a number of white papers and methodologies used across the country. These included:

- The criteria in the California Public Utilities Commission's (CPUC) Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approaches
- The CPUC methodology developed for residential and small commercial customers
- The CPUC methodology developed for large commercial and industrial customers

[2] "Standardized Methods for Free-Ridership and Spillover Evaluation—Task 5 Final Report (Revised)", National Grid, NSTAR Electric, Northeast Utilities, Unitil, Cape Light Compact, June 16, 2003.

- The 2009 white paper developed by Lisa Skumatz for the CIEE, which examined different methodologies for estimating free-ridership
- The methodology used by Wisconsin Focus on Energy for C&I customers
- The decision framework developed for Focus on Energy
- Market research conducted in the Northwest to track the progress of market transformation programs
- Energy Trust of Oregon Fast Feedback Pilot report on collecting rapid feedback from program participants
- The KEMA New Jersey report, which examined acceleration approaches used in different jurisdictions
- Model Energy-efficiency Program Impact Evaluation Guide—A Resource of the National Action Plan for Energy-efficiency, November 2007. Chapter 5. Calculating Net Energy and Demand Savings.

We also reviewed a number of articles published as part of national energy efficiency conference proceedings. The literature review found that many attribution techniques are based on sound methodologies and are consistent with analytical methods used in the social sciences. Furthermore, many of the criticisms of the self-report approach (SRA) can be alleviated through more careful research design, sampling, survey timing, and question wording.

The reviewed literature presented a number of best practice elements for survey design, data collection, and analytic methods. The literature also points out the importance of making the whole process transparent. The question sequence, scoring algorithms, and handling of inconsistent and/or missing data should be included in the report so that stakeholders can understand how each question and its responses impacts the final estimate. Generally, the methodologies reviewed focused on free-ridership SRA measurement, with much less attention given to SRA spillover techniques.

This review, when compared with the existing 2003 methodology, resulted in the following recommendations:

1. Explore acceleration in more detail by calculating the amount of savings attributable to the program over its lifetime
2. Interview respondents to estimate free-ridership no later than 6 months after installation
3. Revise the survey questions to incorporate probes to rule out alternative hypotheses as some of the question batteries reviewed contained more extensive probing on other rival hypotheses for installing the efficient equipment
4. Test one primary set of questions to produce a free-ridership score, while using consistency questions (as well as other information where relevant for larger projects) to adjust the score
5. Examine different methods for handling “don’t know” responses
6. Employ, where feasible, a preponderance of evidence (or triangulation of results) approach that uses data from multiple sources, especially for large savers and complex decision-making cases
7. Change the existing method for estimating spillover (only ‘like’ spillover) to a less conservative approach that includes savings from “unlike” spillover.

1.5 TASK 3: DEVELOP AND TEST METHODOLOGY TO MEASURE FREE-RIDERSHIP AND SPILLOVER

As part of the third task, we drafted a questionnaire and analysis algorithm, then tested the proposed methodology for the self-report approach. The first test we employed was a recently developed approach to assessing method reliability for self-reported net-to-gross surveys, which team members used successfully in the recent California evaluations. The team created a simulated data set of purchase scenarios. One set of researchers specify the scenarios and a subjective range of free-ridership associated with each scenario. A second set of researchers are given the scenarios and independently respond to the survey questions for each scenario. Finally, an analysis team applies the specified net-to-gross calculation algorithm to the survey responses. We then compared the results from the first set of researchers that specified the scenario and provided a subjective range of free-ridership with the results from the analysis team. The results of this testing were used to refine the survey instrument and the analysis algorithm.

The second test was to pre-test the survey instrument with a small number of program participants. As part of this pre-test, the team completed 82 interviews with 2010 participants in National Grid's Small Business Services and Energy Initiative programs. The objectives of the pre-test were to examine whether or not the survey respondents understand the intent of the questions and can respond to the questions. Survey responses were also analyzed using the recommended algorithm. This analysis allowed us to test the reliability of the questions by examining the consistency of responses across questions.

1.6 TASK 4: PRESENTATION OF THE PROPOSED FINAL METHODOLOGY AND FINAL REPORT

Based on the results of the survey pretest, we made minor adjustments to the wording in the pretest survey instruments. The algorithm remained unchanged. The changes to the participant survey included the following:

1. Minor wording changes through the instrument to further clarify question wording. These changes are based on feedback from interviewers and monitoring of pretest interviews.
2. Added prompt for respondents about any interest-free financing they might have received as part of program participation. If the respondent received interest-free financing as part of the program, they are reminded of this assistance during the free-ridership battery.
3. Added prompts for "upgrading equipment" to R4b and R4c.
4. Moved C1 – influence of third party – to follow question regarding third party involvement (FR4).
5. Moved "Impact of Previous Program Participation" section to follow free-ridership battery.
6. Added three questions to assess the program's influence on unlike spillover.
7. Added consistency check scale items to both like and unlike spillover batteries.
8. Added questions regarding expected non-energy impacts.
9. Removed R4b and R4c. These questions were only used for process-related issues as part of the pretesting.
10. Added S1c ("Was this equipment more energy efficient than standard efficiency or code equipment?") to assess partial spillover savings. If the efficiency level of the spillover equipment installed by the respondent was below what would qualify for the program but more efficient than standard or code equipment, the savings are reduced by 50 percent.

To the vendor survey, the following change was made:

1. Added pre and post participation sales volume questions to the vendor survey.

In the final methodology report, we recommend estimating free-ridership in a timeframe close to program participation in order to reduce potential recall bias, and estimating spillover at a later date with these same participants to allow them time to install efficient measures outside the program. For the upcoming free-ridership study of 2010 participants, more frequent measurement for free-ridership purposes is not possible, but should be used in future years.

For the 2011 study of 2010 participants, we have assumed two different data collection periods, one for the C&I electric study with a final report by mid-June and one for the C&I gas study with a final report by August 30, 2011. For the electric study, we have further assumed that we will provide separate free-ridership and early spillover estimates for the following large PAs having a sufficient number of participants—National Grid, NSTAR, WMECo, Unitil, and the Cape Light Compact. For the remaining electric PAs (and in cases where measure categories have less than ten participants), we will apply a statewide average. For the gas study, we will also apply statewide averages where the number of participants and/or measure categories are low.

Based on discussions with the PAs, it was also decided for this survey year to include the spillover questions in the free-ridership customer survey to measure 'early' spillover. The survey will also include a few questions asking about anticipated non-energy impacts from their participation. A more comprehensive spillover effort will occur in the fall and in 2012 as part of the C&I non-energy impact interviews with a subset of the 2010 program participants interviewed in 2011 for the free-ridership and early spillover study. In each annual evaluation cycle, we will focus on measuring the NEI and spillover for a different segment of C&I participants. We will divide the C&I participant population into three categories based upon the type of program, specifically:

- 2011 – C&I Prescriptive Retrofit Programs
- 2012 – C&I New Construction Programs
- 2013 – C&I Custom Programs

Segmenting the customers by program type will allow us to capture the efficiencies in survey design and data analysis on an annual basis while achieving NEI estimates at the conclusion of the three year evaluation period.

2. INTRODUCTION

2.1 INTRODUCTION

The focus of this study is on the general methods for estimating what would have happened absent a program for C&I programs in Massachusetts. The net program effect is the observed effect, less the estimate of what would have happened absent the program. Several methods to estimate what would have happened absent a program are discussed in Chapter 2. The most common method is to estimate free-ridership and spillover.

At this point it is useful to define free-ridership and spillover. Program attribution refers to energy impacts that can be attributed with some level of confidence to program efforts. A program's *free-ridership rate* is the percentage of program savings attributed to free-riders. A *free-rider* refers to a program participant who received an incentive or other assistance through an energy efficiency program who would have adopted the same high-efficiency measure[3] on their own at that same time if the program had not been offered. For free-riders, the program is assumed to have had no influence or only a slight influence on their decision to install or implement the energy efficiency measure. Consequently, none or only some of the energy (and demand) savings from the energy efficiency measures taken by this group of customers should be credited to the energy efficiency program.

In addition to simply identifying free riders, it is important to measure the *extent* of free-ridership for each customer. Pure free-riders (100%) would have adopted exactly the same energy efficiency measure at that time in the absence of the program. Partial free-riders (1–99%) are those customers who would have adopted some measure at that same time on their own, but of a lesser efficiency or a lesser quantity, or they would not have adopted the efficient measures until a later time. Thus, the program had some impact on their decision. Non-free-riders (0%) are those who would not have installed or implemented any program energy efficiency measure (within a specified period of time) absent the program services.

Spillover refers to additional energy efficiency measures adopted by a customer due to program influences, but without any financial or technical assistance from the program. *Participant "like" spillover* refers to the situation where a customer installed energy efficiency measures through the program, and then installed additional measures of the same type due to program influences. *Participant "unlike" spillover* is where the customer installs other types of energy efficient measures than those offered through the program, but are influenced by the program to do so.

Free-drivers, or nonparticipant spillover, refers to any energy efficient measures adopted by program nonparticipants due to the program's influence. The program can have an influence on design professionals and vendors as well as an influence on product availability or practices, product or practice acceptance, customer expectations, and other market effects. All of these may induce nonparticipants to take energy efficiency measures. *Nonparticipant "like" spillover* refers to additional measures of the same type as offered through the program that are adopted due to the program's influence.

Over the past decade, there has been extensive debate across the country regarding the need for and measurement of estimating what would have happened absent a program. Isolating the effects of these

[3] For purposes of this discussion, an "energy efficiency measure" includes high efficiency equipment or appliances, an efficiency measure such as weatherization, or an energy efficient practice such as turning off a computer when not in use.

program factors, and other influences in the decision to adopt energy efficiency measures, is often referred to as "attribution." The increasing importance of program attribution given aggressive savings targets and DSM incentive mechanisms has intensified this debate. Some argue that free-ridership and spillover cancel each other out and should not be measured, that they are too difficult to reliably estimate, or that funds are better spent for program implementation. However, other evaluators and regulators note the advantages of consistent measurement of free-ridership and spillover. Fagan, Messenger, Rufo and Lai (2009)[4], list the following reasons why understanding a program's net savings is important:

- Understand program and portfolio cost-effectiveness
- Improve portfolio design and resource allocation
- Refine program design
- Understand market transformation
- Align program administrators' financial interests with societal interests
- Understand how energy efficiency programs affect baseline load forecasts and short-term power procurement decisions.

2.2 BACKGROUND

Prior to 2003, the Massachusetts Program Administrators (PAs) independently quantified the impacts of free-ridership and spillover (net-to-gross factors) for their energy efficiency programs targeting commercial and industrial (C&I) customers in Massachusetts. These independent evaluation approaches resulted in reported impacts that were measured using a variety of survey instruments, analysis techniques, and assumptions.

In 2003, a consortium of Massachusetts PAs funded a study to develop standardized sampling techniques, data collection approaches, survey questions, survey instrument(s), and an analysis methodology that each of the PAs could use to determine free-ridership and spillover factors for C&I programs.[5] This standardization, which relied on customer and vendor self-reports, was important to the study sponsors for achieving consistency in the annual program results they report to state regulatory agencies. The 2003 final report recommendations for a standardized approach were developed by:

1. Conducting in-depth interviews with each of the five PAs to gain a thorough understanding of their programs and to assess similarities and differences among the programs, measures, and reporting needs,
2. Developing common survey questions, instruments, and an analytical methodology for presentation to the PAs,
3. Moderating a working group meeting with the PAs and NUP consultants to discuss any issues or concerns with the proposed survey questions, instruments, and analytical methodology, and

[4] "A Meta-Analysis of Net to Gross Estimates in California". Jennifer Faga, Mike Messenger, Mike Rufo, Peter Lai, paper presented at the 2009 AESP conference.

[5] "Standardized Methods for Free-Ridership and Spillover Evaluation—Task 5 Final Report (Revised)", National Grid, NSTAR Electric, Northeast Utilities, Unitil, Cape Light Compact, June 16, 2003.

4. Revising the proposed methodology, based on the working group meeting and subsequent comments received from PAs and NUP consultants.

2.3 STUDY OBJECTIVES

This methodology study revisited the 2003 standardized methodology for C&I programs by reviewing other methodologies being used across the nation. This review explores the pros and cons of alternative methods for estimating what would have happened in the absence of the program in different contexts. While the objective of this current study is to develop a standardized methodology for situations where end-users are able to report on program impacts via self-report methods, we recognize that some programs cannot be addressed using a standardized methodology. Therefore, as part of this study we provide a decision framework and guidelines for when the standardized self-report methodology is appropriate and when other methods need to be used (e.g., upstream programs).

2.4 SCOPE OF THE RESEARCH

In order to develop the standardized methodology and the decision framework and guidelines for C&I programs, the study team (Tetra Tech, KEMA, NMR) conducted the four tasks listed below.

Task 1: Kick-off meeting and program review
Task 2: Best practice review
Task 3: Develop and test methodology to measure free-ridership and spillover
Task 4: Presentation of the proposed final methodology and final report

2.4.1 Task 1: Kick-off meeting and program review

At the kickoff meeting with the key contacts from each PA, the study team discussed the objectives for the meeting, presented an overview of the existing 2003 methodology and its limitations (e.g., reliance on self-reports, generous handling of acceleration, conservative spillover measurement, limited market effects), discussed potential issues for consideration in revising the methodology for C&I programs (e.g., changes in behavior, changes in O&M practices), and discussed the four tasks to be completed as part of this study.

As part of the program review, the study team reviewed the three-year plans and information already collected from the PAs by the KEMA team as part of the current C&I evaluations being conducted under separate contracts. Where additional information needed to be collected to develop a complete understanding of the programs (or specific measures within programs), and to develop the decision framework and methodology, the cross-cutting study team attempted to re-contact each PA to ask additional questions. The product of this review is a taxonomy of programs by sector, type of assistance, eligibility, incentives, and delivery. It can be used to determine the type and number of approaches for review and development (Appendix A provides this taxonomy).

2.4.2 Task 2: Best practice review

The purpose of the best practice review was to identify and classify variations of net-to-gross methodologies to inform the recommended methodology and the decision framework. As part of the best practice review, we reviewed and synthesized recent attribution research and compared it to the existing standardized methodology. We also reviewed a number of white papers and methodologies used across the country (see Appendix B for a complete list of articles and approaches reviewed). These included:

- The criteria in the California Public Utilities Commission's (CPUC) *Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approaches.*

- The CPUC methodology developed for residential and small commercial customers.
- The CPUC methodology developed for large commercial and industrial customers.
- The 2009 white paper developed by Lisa Skumatz for the CIEE, which examined different methodologies for estimating free-ridership.
- The methodology used by Wisconsin Focus on Energy for C&I customers.
- The decision framework developed for Focus on Energy.
- Market research conducted in the Northwest to track the progress of market transformation programs
- Energy Trust of Oregon Fast Feedback Pilot report on collecting rapid feedback from program participants
- The KEMA New Jersey report which examined acceleration approaches used in different jurisdictions.
- Model Energy Efficiency Program Impact Evaluation Guide—A Resource of the National Action Plan for Energy Efficiency, November 2007. Chapter 5. Calculating Net Energy and Demand Savings.

The results of the review are presented in Chapter 3 of this report.

The final steps in Task 2 of this methodology study were to conduct a web meeting with the PAs to discuss the recommendations and agree on approach(es), and finalize this report on best practices, recommended methods, and a decision framework.

2.4.3 Task 3: Develop and test methodology to measure free-ridership and spillover

Based on our review of these documents and our compilation of a taxonomy of C&I programs offered in Massachusetts, in Task 3 we developed and tested the methodology to measure free-ridership and spillover. The results of this test and our final recommendations are presented in Chapter 4 of this report.

2.4.4 Task 4: Presentation of the proposed final methodology and final report

Task 4 summarizes our final proposed methodology. This methodology will be used in the 2011 free ridership and spillover studies with 2010 participants.

2.5 OVERVIEW OF REPORT

Chapter 2 discusses the differing methodologies available for studying what would have happened in the absence of the program, while Chapter 3 presents the findings from the literature review on self-report methodologies. Chapter 4 presents the pretest results from the revised methodology, and Chapter 5 presents our recommendations for the methodology for the full scale free-ridership and spillover study of 2010 participants.

The taxonomy of PA programs by sector, type of assistance, eligibility, incentives and delivery is contained in Appendix A. Appendix B lists the sources reviewed as part of the best practice review, Appendix C summarizes articles reviewed, Appendix D contains the results of mean and median substitution testing from the pretest, Appendix E contains the final revised participant and vendor survey instruments (and a summary of changes made to the instruments based on the pretest), and Appendix F contains the final algorithm flowcharts.

3. METHODOLOGIES FOR STUDYING FREE-RIDERSHIP AND SPILLOVER

3.1 TYPES OF METHODOLOGIES AND THEIR ADVANTAGES/DISADVANTAGES

3.1.1 Overview

The method to estimate what would have happened absent a program can vary by program type. In some cases, method selection depends on the details of program and may be a matter of judgment or degree. Others methods are fundamentally not applicable for certain program categories. The most common methodology is self-report by participating customers. Other methods, discussed further below in Section 2.1.2 include market sales data analysis, retailer or contractor surveys, pricing and elasticity analysis, structured expert judging, historical tracing (case study method), billing analysis, estimating spillover, program delivery staff surveys, and on-site data collection/M&V contractors.

Self-reports by participating customers is the approach historically used to estimate free-ridership and spillover in Massachusetts. It can be used with most any type of downstream program. The self report methodology asks customers about the effect of the program on their decision to adopt specific measures. The self-report approach has often been criticized. For example, Peters and MacRae[6] identify the following primary concerns:

- Potential biases related to respondents' giving "socially desirable" answers,
- The inability of customers to know what they would have done in a hypothetical alternative situation, especially in current program designs that use multiple methods to influence behavior
- The tendency of respondents to rationalize past decisions on choices
- Potential arbitrariness of scoring methods that translate responses into free-rider estimates
- Lack of customer recognition of the influence the program may have on other parties influencing their decisions (e.g., program may have influenced contractor practices, which in turn may indirectly impact the participant's decision).

While these concerns all have validity, they have been recognized for as long as free-ridership has been estimated. Good survey techniques can mitigate many of these potential biases. (See, for example, Ridge et al.[7]) Moreover, the majority of methods available for estimating program effects on customer measure adoption are based on some type of surveys. Even market sales data often come from survey responses which may be incomplete or subject to various biases.

---

[6] Free-ridership Measurement Is Out of Sync with Program Logic…or, We've Got the Structure Built, but What's Its Foundation?, Jane S. Peters and Marjorie McRae, Research Into Action, Inc. In Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings. Washington, DC., 2008.

[7] The Origins of the Misunderstood and Occasionally Maligned Self-Report Approach to Estimating the Net-To-Gross Ratio, Ridge, Richard, Ridge & Associates, Philippus Willems, PWP Inc, and Jennifer Fagan, Itron, Inc., and Randazzo, Katherine, KVD Research Consulting. Presented at Proceedings of the 2009 International Energy Program Evaluation Conference, Portland, Oregon.

Thus, while there are certain biases working toward over-reporting free-ridership, there are others working toward under-reporting free-ridership. Some of the recent literature paints a bleak picture of our ability to learn anything by talking to people. However, there are well established methods that can mitigate many of these problems. In particular, biases in both directions can be mitigated by well designed surveys using good set-up questions. Furthermore, scoring systems can be validated and calibrated by methods like what was done for the residential/small customer net-to-gross protocol for the California Public Utilities Commission 2006-08 evaluations[8].

In addition, the self-report survey methods, particularly for larger commercial and industrial customers, may be enhanced by using other data sources to identify the level of free-ridership for a particular project. For example, on-site data collection, personal interviews at the site, or project file reviews for impact evaluation may provide further evidence to confirm or refute the self-reports of factors used to determine free-ridership or spillover. Other market intelligence includes whether energy efficiency measures were readily available in the absence of the program. In any event, adjustments to project or program free-ridership need to be clearly documented to show the findings are objective and not subject to any bias by the evaluator or M&V contractor.

Thus, in developing a typology of free-ridership and spillover methods, we do not focus on whether or not a "self-report" methodology is used. Instead, we identify who the respondents or data providers are, what types of information they are providing, and how the data are analyzed. For each method described in these terms, we then indicate key factors affecting the method's validity, and key data collection issues.

These basic characterizations are provided in Section 2.1.3. We consider how well each method would apply to measures with particular features. We also give a qualitative indication of the cost or complexity level.

A key factor in the validity of any survey-based method is good survey design. In section 3, we describe features that strengthen the validity of survey-based net-to-gross (NTG) methods. In comparing methods, we assume that these validity-enhancing features would be included in all methods where applicable.

Methodology characterizations, in terms of risks to validity, applicability, and costs, are never black and white. As a result, the choice of the "best" method for a particular situation is not clear-cut. In the remainder of this chapter, we describe some general methods, and provide summary tables of the key features of specific methods. In Section 2.2, we provide a decision framework for choosing among alternative methods for particular contexts.

3.1.2 Discussion of general approaches

a. Market sales data analysis

Approaches based on market sales data analysis are in many ways the ideal when the necessary data can be obtained. These methods capture the total net effect of the program, including both free-ridership and participant and nonparticipant "like" spillover. However, this ideal is often not possible to implement.

---

[8] Response to Overarching Comments Regarding the Use of Self-Reported Net-to-Gross (NTG) and the Residential and Small Commercial Self-Report Approach NTG Method, January 28, 2010, Residential/Small Commercial Joint Simple Net-to-Gross (Self-Report) Committee.

The challenge of this method is in obtaining comprehensive market data for both the area of interest and an appropriate comparison area, and/or having an appropriate non-program baseline for the study area. The "baseline" means an estimate of the sales that would have occurred in the program area absent the program. Ideally, program effects are estimated not just by comparing pre-program data to with-program data from the program area, nor by comparing the program area with a comparison area. Rather, the change in the program area is compared with the change in the nonprogram area.

It is much more common for overall efficient market share to be compared in the program area versus the comparison area, and to attribute the observed differences to the program, subject to a good faith effort to confirm or disconfirm the comparability of the two areas. The problem with using a pre-post/test-comparison design is that long-standing programs tend to render the program and comparison areas systematically non-comparable in ways that are aggravated by this particular experimental design.

It is not necessary that the comparison area be exactly comparable to the area of study interest. It is necessary that a credible baseline for the area of interest can be constructed based on the comparison-area data, possibly with a set of systematic adjustments. For example, sales data may be expressed in terms of sales shares to control for differences in total size of the two areas; if available, shares may be calculated separately by segment to allow adjustment for a different segment mix between the areas.

b. Retailer or contractor surveys

Retailer or contractor surveys can be used in several key ways for NTG estimation.

1. For custom or relatively rare measures, the vendor can be asked about the decision process and influence for individual customers. This information can be combined with the customer's report of influences. If either the customer or the vendor reports that the program influenced the purchase/or sale, some amount of credit is given to the program depending upon the degree of influence.

2. For high-volume measures, the vendor is asked about general influences of the program on the vendor's level of sales and stocking patterns. This general information can also be combined with customer-reported influence of the vendor and program on the purchase decision.

3. Vendors can be asked about sales volumes and efficient equipment sales shares for conditions with and without the program, or for in-territory and comparison area sales. This approach can be analyzed similarly to market-level sales data. The difference is that the market sales data approach usually refers to comprehensive or nearly comprehensive reporting of sales. By contrast, vendor surveys may collect "best guess" estimates of sales volumes and shares from a sample, then use sampling weights and other measures of size (such as employment) to expand the survey responses to the full market. This is the approach being taken for the Massachusetts C&I New Construction High Bay Lighting Market Effects study.

Either of the first two approaches uses a combination of customer and upstream actor surveys to capture the effect of the program on the upstream actors, who in turn influence the end-user. Such studies, which combined participating customer and participating vendor self-reports in this way, found that:

- The end-user survey indicates the influence of the incentive or other assistance the customer received, as well as the influence of the vendor or other supply-side agent, in the customer's decision to adopt a measure

- The supplier survey indicates the role of the program in changing the supplier's promotional efforts

- Together these two sources indicate the likely role of the program in the customer's decisions to adopt the measure.

c. Pricing and elasticity analysis

Several methods can be used to estimate price elasticity; that is, the effect on purchases by lowering the price through upstream or downstream incentives.

Stated Preference experiments systematically ask potential customers what they would choose from a set of options with different features and prices. The "choice sets" offered each customer are designed so that the effect of price and features can be estimated from the data set of all the customers' responses. Conjoint and double bounded techniques are two approaches structured for this type of estimation.[9] The key challenge for stated preference methods is the potential difference between (1) what customers say they will buy in a hypothetical situation, and (2) what they do buy when they actually have to spend money.

Simpler stated preference surveys can be used when features are not being investigated and the only question is how a change in price affects purchases.

Revealed Preference studies observe the actual choices customers make from true choices available to them when making purchases. The key challenges for revealed preference studies are:

- It is necessary to observe choices in contexts that are similar to those of the study area, except for the presence of the program. This is a similar challenge to that of obtaining a valid comparison area for market sales data. The difference is that if the valid comparison area can be defined, it may be more practical to collect revealed preference data there than to obtain comprehensive sales data.

- To obtain accurate revealed preference information, it is usually necessary to observe the items purchased. Customers cannot reliably report the efficiency levels of recently purchased equipment. Direct observation can be accomplished via store intercepts for small items such as light bulbs, or via onsite visits for large items such as refrigerators. The remaining challenge for this method is the potential nonresponse bias; that is, potential differences between customers willing to have their purchases observed and those who decline.

To obtain a NTG estimate from the survey-based elasticity estimates, it is necessary to estimate the effect of the program on prices in the region. The price effects can be estimated using shelf surveys. Alternatively, price effects can be asked about in supplier surveys at various levels.

---

[9] Conjoint experiments provide customers with a series of hypothetical choices among a set of products with varying combinations of attributes and varying prices. Double bounded questions ask a series of questions to determine upper and lower bounds on the price at which a customer would buy a particular product.

d. Structured expert judging

Structured expert judgment studies assemble panels of individuals with close working knowledge of the technology, infrastructure systems, markets, and political environments addressed by a given energy efficiency measure to estimate baseline market share and, in some cases, forecast market share with and without the program in place. Structured expert judgment processes employ a variety of specific techniques to ensure that the participating experts specify and take into account key known facts about the program, the technologies supported, and the development of other influence factors over time.

The *Delphi process* is the most widely known method of this family of methods. Properly executed, a Delphi process includes at least one iterative step where each expert is provided with the judgments and rationales of all the other experts, and offered an opportunity to re-assess or defend the original position.

In the context of energy efficiency program evaluation, structured expert judging has as its foundation the experts' experience with other programs and the NTG findings for other programs, typically based on other methods. Structured expert judging allows experience from other contexts to be applied to situations where the same methods may have more threats to validity. Expert judging also allows adjustments to be made, albeit subjectively, for some of these threats.

A particularly useful role for structured expert judging is to develop a "consensus" estimate to consolidate results from multiple estimation methods. In the context of the Massachusetts programs, this is the role we recommend for this method. We do not recommend using Delphi or related methods as stand-alone estimates.

e. Historical Tracing: Case Study Method

This method involves the careful reconstruction of events leading to the outcome of interest, for example, the launch of a product or the passage of legislation, to develop a 'weight of evidence' conclusion regarding the specific influence or role of the program in question on the outcome. Historical tracing relies on logical devices typically found in historical studies, journalism, and legal argument. These include:

- Compiling, comparing, and weighing the merits of narratives of the same set of events provided by individuals with different points of view and interests in the outcome.
- Compiling detailed chronological narratives of the events in question to validate hypotheses regarding patterns of influence.
- Positing a number of alternative causal hypotheses and examining their consistency with the narrative fact pattern.
- Assessing the consistency of the observed fact pattern with linkages predicted by the program logic model.

Researchers use information from a wide range of sources to inform historical tracing analyses. These include public and private documents, personal interviews, and surveys.

This method is best suited to attribution analysis of major events such as adoption of new building codes or policies. It is not typically applicable to energy efficiency programs, and is not recommended as a primary method for the Massachusetts programs. However, elements of this approach may be used in analysis of very large custom projects that require essentially case study approaches.

f. Billing Analysis

Billing analysis estimates program savings by analyzing consumption data. The most common form of such analysis compares usage after program participation with usage prior to program participation, with some form of weather normalization. Analysis may be done separately for each customer and then aggregated, or may be done in a pooled time series cross-sectional model. Since billing analysis typically requires up to 12 months of post-implementation consumption data, this approach may not be feasible where more timely feedback on net savings is required.

When a comparison group is used in the analysis, the resulting estimate of savings is sometimes interpreted as savings net of free-ridership. The rationale is that the comparison group change in usage represents how the participants would have changed in the absence of the program.

This approach depends heavily on the "comparability" of the comparison group. In most programs, customers who choose to participate are different from those who do not participate in ways that can affect their year to year consumption changes. This self-selection effect can be controlled for, to some extent, by including in the analysis customer characteristics that might be associated with both the propensity to participate and the consumption changes absent the program.

The most complete treatment along these lines is the Inverse Mills ratio self-selection correction developed by J. J. Heckman. Ken Train and Miriam Goldberg expanded this approach to the context of Statistically Adjusted Engineering analysis for program evaluation. However, this approach still assumes that there is no correlation between the change in usage not explained by the model and the remaining unobserved factors that determine program participation. This assumption is difficult to justify.

The assumption that the comparison group change is a good representation of how participants would have changed absent the program can be justified in some cases. A common example is low-income programs, where past and future participants may serve as a comparison for current participants. In the commercial/industrial context, the participants are usually too diverse, and there are too many non-program changes, to get stable, reliable estimates from aggregate billing analysis.

In cases where there is a credible comparison group and model or analysis structure, the billing analysis provides net savings. The analysis does not isolate a net-to-gross effect from adjustments to gross savings.

A second situation where a valid comparison is available occurs when customers are randomly assigned to receive the participant "treatment" or not. With random assignment, there is no systematic difference between the untreated or control group customers and the participating customers, other than the treatment itself. Therefore, the control group provides an unbiased estimate of what participants would have done in the absence of the program treatment.

This approach is rarely possible in full-scale programs, but is sometimes possible in pilots. The OPower program that has been instituted in several territories in recent years relies on this type of random assignment and comparison of consumption data to determine savings from an informational treatment. Other behavioral programs using this type of approach are under development or consideration.

The random assignment informational programs use many thousands of customers in the treated and control groups, far more than in typical pilots. On the other hand, the ability to estimate savings from these programs requires the use of random assignment, unlike a typical full-scale program.

For any type of program, if the comparison group is otherwise valid, but there is nonparticipant spillover, net savings estimation based on comparison of participants with nonparticipants will understate the savings attributable to the program. Essentially, the nonparticipant spillover will be counted as part of naturally occurring savings, and subtracted from participant change. Thus, not only is the program not credited for the nonparticipant spillover, but also the true participant savings attributable to the program is underestimated.

In the case of an informational program with random assignment to treatment or control, there is a possibility that treated customers may influence their control-group neighbors to take actions they otherwise would not. This influence would be a case of nonparticipant spillover from the program. The program savings estimated via billing analysis would not include the nonparticipant spillover; additionally, the participant savings would understate the amount of the nonparticipant spillover.

Another effect included in the net savings yielded by billing analysis is "takeback". Takeback is an increase in usage after an energy efficiency improvement as a behavioral response to the lower cost of using the equipment after the improvement. Takeback effects arguably should be considered in gross savings, not as net-to-gross. The net-to-gross factor indicates how much of the gross savings occurred due to the program and would not otherwise have occurred. If there is takeback associated with a measure, that takeback in most cases is a characteristic of the measure and not of whether or not the program influenced the measure implementation. For example, if the effect of adding better insulation is that occupants set their thermostats warmer, this is not an effect of the program but an effect of the measure, whether the measure was program-induced or not.

Like nonparticipant spillover, takeback is difficult to isolate from other factors affecting participant and nonparticipant changes. Agnew and Goldberg[10] present analysis exploring the relationship between takeback estimates and engineering assumptions in the context of HVAC measures. Takeback can be measured to some extent via surveys asking what people are doing differently since getting the measure, and to what extent that is because of the measure. This approach means relying again on self-reports.

g. Estimating spillover

Market sales data estimates, described above, provide a combined NTG that includes both free-ridership and participant and nonparticipant "like" spillover. Participating customer surveys can also collect information on like spillover. However, the best timing for when to ask about free-ridership (as close as possible to the time of the program-supported measure) may be too soon to learn about spillover effects on later purchases. Supplier surveys can address spillover, in terms of how the sales of efficient equipment outside the program have been affected by the program.

Whether revealed preference or stated preference surveys capture spillover depends on the structure of the program and on the study design. For an upstream program, these methods will capture the total change in purchase rate due to the price reduction caused by the program, or due to other factors. For a rebate program, it is possible for revealed preference surveys combined with discrete choice analysis to tease out the effect of the rebate both on those who take the rebate and those who do not. On the

[10] Getting to the Right Delta: Adjustment and Decomposition of Billing Analysis Results, Agnew, Ken and Goldberg, Miriam, *Proceedings of the 2009 International Energy Program Evaluation Conference.*

other hand, a stated preference survey for a rebate program is not typically able to make this separation accurately.

h. Program delivery staff surveys

Some practitioners have used reports from program staff on their influence on customers as input to NTG estimates, particularly for custom projects. An argument for this approach is that customers may be inclined to want to take credit for a good idea, and would not necessarily remember how the program staff helped provide information and develop the project. A critical counter-argument is that program staff have an explicit vested interest in obtaining high attribution credit, which the customers do not. Moreover, program staff do not necessarily realize when the suggestion they made was already something the customer was considering.

Thus, we do not recommend using delivery staff reports of their influence on customers as a valid basis for estimating NTG. We do however recommend obtaining information from delivery staff as background for understanding and clarifying projects and decision points.

For all the methods that include customer ex-post counterfactual surveys, our proposed approaches would start by collecting all relevant information the program can provide on the specific projects targeted by the surveys (for custom projects) or on the general operation of the program, for prescriptive measures. This information should include any detail the program can provide on the nature and timing of program information and assistance for the project. For custom projects, the project file may include notes on when the customer first contacted the program staff and what they were considering in terms of energy efficiency measures. While this information does not go into explicit free-ridership scoring or calculations given the qualitative nature of the information, it should be used in probing customer responses and as a consistency check.

A second important feature of our recommendations on survey approaches is that the surveys be designed to minimize potential biases[11] in customer and vendor responses. Starting with good information from the program is an important step in this design. Additional elements of good survey design to mitigate potential biases are discussed in Section 3.1.2 of this report.

i. On-Site Data Collection/M&V Contractors

Another source of information for custom projects may include the contractor conducting on-site data collection for M&V or impact evaluation. They are in a position to determine whether other energy efficiency measures were taken and the timeframe relative to the existing project covered under the program. Further, it may be useful to take a case study approach when there are few projects but they represent large savings. Case studies could include on-site interviews of other customer staff or telephone surveys of multiple design team members.

[11] Examples of potential biases include bias related to respondents giving socially desirable responses, bias related to the tendency of respondents to rationalize past choice decision, and the inability of customers to know what they would have done in a hypothetical alternative situation.

3.1.3 Summary of methods

Survey-based NTG methods can be described in terms of the following:

- Who the respondent is or from whom the data are obtained: participating or nonparticipating customers, participating or nonparticipating suppliers at various points in the supply chain
- What types of information are obtained from the respondent
- How the data are analyzed.

Table 3-1 summarizes the common NTG methods available based on these elements. For each of these methods, Table 3-2 indicates key research design factors affecting method validity, as well as issues that need to be addressed in the data collection. The table also includes methods that are not necessarily survey based, but can be described in similar terms.

Table 3-1. Common NTG Methods

Surveyed Group	Types of data collected	Data Description	Analysis	Includes Like Spillover	Includes Unlike Spillover	Includes Nonparticipant Like Spillover
Participating End-users	Post Hoc counterfactual:	Self-reported likelihood of buying absent program assistance	Scoring and averaging	Optionally	Optionally	No
Participating and nonparticipating end-users	Revealed preferences	Actual purchases, prices, and customer characteristics	Discrete choice analysis or simple average adoption rates	Optionally	No	Optionally
Nonparticipating end-users	Stated preferences	conjoint or double bounded exercises	Conjoint or double bounded specialized analysis	Yes	No	Yes
Nonparticipating end-users	Stated preferences	likelihood of purchase at varying conditions	Scoring and averaging	No	No	No
Manufacturers & Regional buyers and distributors	Market sales data	Sales of efficient and standard equipment in program and non-program areas over time		Yes	No	Yes
Retail store managers and contractors	Sales data	Sales of efficient and standard equipment when program is and isn't present	Weighting and/or averaging	Yes	No	Yes
Retail store managers and contractors	Sales and counterfactual	Promotional activity and sales with and without program	Weighting and/or averaging	Yes	No	Yes
Retail store managers and contractors	Customer-specific influence	Customer-specific influence of program and supplier	Weighting and/or averaging	Typically not	No	No
Retail store	Shelf and stocking observations	Observed shelf volumes and prices	Modeling or averaging	Depends on accompanying elasticity analysis	No	Depends on accompanying elasticity analysis
Participating customers and comparison group customers	Billing data	Consumption data for roughly one year pre- and post-participation	Weather normalization and change analysis	Yes	Yes	Counted negatively
Experts		NTG estimates from multiple methods	Delphi process	Depends on input methods	Depends on input methods	Depends on input methods

3.1.4 Design and data collection considerations

For any research method, details of the situation and of the study design affect the method's validity. The table below indicates design factors and data collection issues that must be considered in implementing each of the NTG methods listed in the previous table. Some of these factors are within the control of the researcher, whereas others cannot be controlled.

Table 3-2. Design Factors and Data Collection Issues to be Addressed for Common NTG Methods

Surveyed Group	Types of data collected	Validity: Depends on—	Data collection issues
Participating end-users	Post Hoc counterfactual:	scoring; accuracy of self-reported hypothetical action absent the program	Requires well designed surveys to mitigate response bias.
Participating and nonparticipating end-users	Revealed preferences	comparable conditions between purchases in presence of program and those in absence of program; accuracy of purchase and pricing data obtained from customers	Typically requires in-store intercept or onsite observation
Nonparticipating end-users	Stated preferences	Match of stated preferences for hypothetical purchases to actual experience; customers in exercise see and consider information more systematically than in actual purchases	Requires recruitment of customers willing to engage in extended exercise. Requires companion study to determine change in prices and stocking absent the program
Nonparticipating end-users	Stated preferences	scoring; accuracy of self-reported hypothetical actions if purchasing with and without program	Requires companion study to determine change in prices and stocking absent the program
Manufacturers & Regional buyers and distributors	Market sales data	ability to construct non-program baseline from pre-program and/or comparison area market coverage	Often some key suppliers don't cooperate, "holes" need to be plugged
Retail store managers and contractors	Sales data	market coverage and weighting; accuracy of supplier recall if not based on sales records	Often some key suppliers don't cooperate, "holes" need to be plugged
Retail store managers and contractors	Sales and counterfactual	market coverage and weighting; accuracy of supplier-reported actual and hypothetical activities	
Retail store managers and contractors	Customer-specific influence	accuracy of supplier's report on influencing factors for individual customers	
Retail store	Shelf and stocking observations	stores that represent a meaningful comparison set to those with active programs	Need access to stores; some companies deny access
Participating customers and comparison group customers	Billing data	Valid comparison group with minimal self-selection effects.	
Experts		Well documented methods Effective iteration process Experts' expertise	

3.2 DECISION FRAMEWORK FOR SELECTING APPROPRIATE METHODOLOGIES

3.2.1 General

What NTG methodology makes the most sense for a given program and measure depends on the structure of the program, as well as the types of data that can realistically be obtained. Thus, key dimensions to be considered in choosing a method are the following.

a. Availability of market sales data, with meaningful comparison group

If market sales data are available indicating the total sales of efficient and standard equipment, AND if these data are available for the program area and time period and also for a meaningful comparison area and/or time period, total program effects may be estimated based on these data. If the data for the program area are unavailable, or if there is no good or meaningful comparison area, group, or time period, this method cannot be used.

b. Homogeneity of the measure and the customers

Most of the available methods require large numbers of similar measures and similar customer types. For custom measures, measures with few participants, or measures with applications in widely disparate conditions, methods based on market data or on samples of customers making similar purchase decisions do not easily apply. The only methods that work well for custom or case-specific measures are end-user post-hoc counterfactual surveys, and vendor surveys asking about specific customers. Since custom programs are likely to have fewer projects, one very large project can have a significant influence on the free-ridership or spillover rate. Therefore, it is worth including multiple approaches, if possible, to confirm the findings.

c. Likelihood of substantial upstream effects unknown to end-use participant

When there is a reasonable likelihood of substantial upstream effects that an end-use participant would not know about, participating end-user counterfactual surveys alone will understate the effect of the program, even if the customer answers accurately from the customer's own perspective. For example, the participating customer would not know that the program influence has changed what options are available, lowered the price of the efficient options, or increased the sales staff's knowledge of and interest in promoting the efficient option. These situations either require information for the market as a whole, if the market sales based approach is viable, or else require a combination of participant end-user and vendor surveys.

d. Cost/Value Tradeoffs

Some methods can provide more credible results, but are also more costly. This cost may be justified for program components that are important to the portfolio, but not for other components. Importance to the portfolio is typically related to the level of spending and/or savings associated with a program component, but may also depend on future program plans, or other "visibility" factors. In the tables below, we indicate rough levels of cost or difficulty typical for each method, but these can vary with individual circumstances.

3.2.2 Methods applicable for different conditions

The table below summarizes which methods are suitable for programs with particular features. The table does not attempt to prescribe a best method for a given situation. Rather, it indicates what options will do better or worse for programs with the indicated features. For a particular program in a

particular context, the choice of methods can be made by balancing the advantages and disadvantages of each.

Table 3-3. Summary of methods applicable to different conditions

Surveyed Group	Types of data collected	Applicability				Typical Cost or Complexity
		Custom Measures	Measures with few, diverse participants	Prescriptive Measures with large numbers of similar participants	Measures with substantial upstream influence invisible to customers	
None	Comprehensive market sales data for program area and comparison area	Poor	Poor	Good	Good	Low if data are available, High or not possible if data need to be developed
Participating End-users	Post Hoc counterfactual:	Good	Good	Good	Poor unless combined with retailer or contractor surveys	Medium
Participating and nonparticipating end-users	Revealed preferences	Poor	Poor	Good	Good if combined with pricing study	High
Nonparticipating end-users	Stated preferences—experiment	Poor	Poor	Good	Good if combined with pricing study	High
Nonparticipating end-users	Stated preferences—likelihood	Poor	Poor	Good	Good if combined with pricing study	High
Manufacturers & Regional buyers and distributors	Market sales data	Poor	Poor	Good	Good	Low
Retail store managers and contractors	Sales data	Poor	Poor	Good	Good	Medium
Retail store managers and contractors	Sales and counterfactual	Poor	Poor	Good	Good	Low
Retail store managers and contractors	Customer-specific influence	Good	Good	Medium	Good	Medium
Retail store	Shelf and stocking observations	Poor	Poor	Good if combined with stated or revealed preference study	Good if combined with stated or revealed preference study	High
Participating customers and comparison group customers	Billing data	Poor	Poor	Good if have valid comparison group	Good if have valid comparison group	Low
Experts		Depends on quality of input methods				Low

3.2.3 Application of decision framework to Massachusetts C&I programs

The application of immediate interest for the NTG decision framework outlined above is the Massachusetts C&I programs. For these programs, market sales data is unlikely to be available for most measures. Econometric methods involving discrete choice analysis of revealed or stated preference data do not work well for custom measures or for measures with relatively few participants. Thus, for these programs, the primary recommendation is to rely on participating end-users plus design team member interviews, and supplier surveys. For the very largest custom projects, project file reviews and program staff interviews may also be used to enhance the analysis.

For program components with little likely effect on the broader market, well designed participating customer surveys may be adequate. For components expected to influence the market non-trivially, beyond the effects on individual participants, the end-user surveys should be combined with supplier surveys. The supplier survey would collect data on sales volumes and shares with and without the program. This can be done by asking suppliers about current and prior practice, and/or about current sales and hypothetical sales if the program did not exist; alternatively suppliers in a comparison area can be surveyed.

Table 3-4 presents the suggested approach by program component type.

Table 3-4. Recommended Approaches for Massachusetts C&I Programs

Program	Importance to Portfolio	Measure Homogeneity	Customer Homogeneity	Likelihood of important influences unknown to customer	Likelihood of large NP effects	Market data availability	Availability of meaningful comparison area or non-program market baseline	Recommended Methods				
Large C&I retrofit	High	Custom and/or few	Low	Medium	Low	Low	N/A	Participating End-users including Design Team Members	Post Hoc counterfactual:			
Large C&I Retrofit	High	Prescriptive, large numbers	Low	Medium	Medium	Low	Possible	Participating End-users	Post Hoc counterfactual:	+	Retail store managers and contractors	Sales and counterfactual
Small C&I retrofit	Medium	Prescriptive but few	Medium	Medium	Small	None	None	Participating End-users	Post Hoc counterfactual:	+	Retail store managers and contractors	Customer-specific influence
Small C&I retrofit	Medium	Prescriptive, large numbers	Medium	Medium	Medium	Possibly for smaller gas measures	Possible	Participating End-users	Post Hoc counterfactual:	+	Retail store managers and contractors	Sales data
C&I New Construction/Major Renovation	Medium	Custom and/or few	Low	Low	Low	Low	N/A	Participating End-users	Post Hoc counterfactual:			
C&I New Construction/Major Renovation	Medium	Prescriptive, large numbers	Medium	Low	Medium	Low	N/A	Participating End-users	Post Hoc counterfactual:	+	Retail store managers and contractors	Sales data

4. LITERATURE REVIEW ON SELF-REPORT APPROACH (SRA) METHODOLOGIES

Since the objective of this study was to develop a standardized methodology for situations where end-users are able to report on program impacts via self-report methods, much of the best practice review focused on those methodologies for C&I programs. Since small commercial methodologies are often appropriate for residential customers, much of this methodology could be applied to residential programs in the future. However, before doing this, more information would need to be collected on the residential programs to determine what parts of the methodology can be shared. This internal program review on residential programs is being conducted in early 2011.

4.1 BEST PRACTICES IN SELF-REPORT APPROACH (SRA) METHODOLOGIES

Free-rider and spillover estimates inherently rely on counterfactuals: what would or would not have happened absent a program. Every method of estimating the counterfactual relies on certain assumptions, as well as on data collection. Which methods of data collection and analysis can produce the most reliable, meaningful results can vary with delivery method and market conditions. In some states (e.g., Wisconsin and California), the choice of net-to-gross estimation methods is guided by a decision framework that considers factors including market sales data availability and quality, ability of respondents to provide accurate information, the likelihood of large nonparticipant effects, and the uniformity of the measure. The decision framework outlined in Section 2 considers similar factors.

The literature review found that many attribution techniques are based on sound methodologies and are consistent with analytical methods used in the social sciences[12, 13] Ridge, Willems, Fagan, and Randazzo (2009) point out that it does not make sense to compare all self-report approaches equally, as some conform to best practice, while others do not.[14] Keating (2009) adds that many of the criticisms of the SRA can be alleviated through more careful research design, sampling, survey timing, and question wording.[15]

4.1.1 Necessary data elements for implementation of SRA free-ridership and spillover measurement

In order to employ best practice methods, it is critical that the PAs (or program implementers) keep complete and accurate electronic records of data needed to implement the SRA approach. This includes, but is not limited to, contact information on customer decision makers, including design team members for custom projects; contact information for contractors and vendors involved with the project; detailed information on services provided (e.g., type of technical assistance), measure or services incentivized

---

12 "Self-report Methods for Estimating Net-to-gross Ratios in California: Honest!", Richard Ridge, Philippus Willems, Jennifer Fagan, paper presented at AESP national conference, San Diego, CA, January 27-29, 2009.

13 "Response to Overarching Comments Regarding the Use of Self-Reported Net-to-Gross (NTG) and the Residential and Small Commercial Self-Report Approach NTG Method", paper presented to the California Public Utilities Commission by Members from Residential/Small Commercial Joint Simple NTG (Self-Report) Committee, the Large Nonresidential NTG Committee, and the evaluation contractors, January 28, 2010.

14 "The Origins of the Misunderstood and Occasionally Maligned Self-Report Approach to Estimating Net-to-Gross Ratio, Richard Ridge, Phillipus Willems, Jennifer Fagan, Katherine Randazzo", paper presented at the 2009 Energy Program Evaluation Conference, Portland.

15 "Free-ridership Borscht: Don't Salt the Soup". Ken Keating, paper presented at the 2009 IEPEC conference.

through the program; rebates paid (at the measure level); gross energy and demand savings (at the measure level); date of installation; and project cost. In addition, there should be a unique identifier for each project and a way to link individual measures to those projects.

In past free-ridership studies that Tetra Tech staff (as PA Consulting Group) have conducted for PAs in Massachusetts (as well as utilities in many other states), data quality and completeness have been an issue. For some, we have been unable to contact program participants and influential vendors due to a lack of contact names or phone numbers. In other cases, total project cost figures are unavailable, making it difficult to remind the participant of the total impact of program participation.

4.1.2 Elements of good design for self-report free-ridership and spillover measurement

The reviewed literature presented a number of best practice elements for survey design, data collection, and analytic methods. The literature also points out the importance of making the whole process transparent. The question sequence, scoring algorithms, and handling of inconsistent and/or missing data should be included in the report so that stakeholders can understand how each question and its responses impacts the final estimate.

Generally, the methodologies reviewed focused on free-ridership SRA measurement, with much less attention given to SRA spillover techniques.

a. Survey design elements

Important survey design elements prevalent in the literature for estimating free-ridership (and spillover) include:

- Identifying the key decision-maker(s) for the specific project. This may include one or more staff at the customer site, including design team members and other market actors (e.g., architects, vendors, trade allies) who were influential in the decision-making process.
- Use of set-up or warm-up questions to help the decision maker(s) recall past events, the sequence of these events, and how they events affected their decision to adopt the measure.
- Use of multiple questions to limit the potential for misunderstanding or the influence of individual anomalous responses.
- Use of questions to rule out rival hypotheses for installing the efficient equipment.
- Testing questions for validity and reliability.
- Setting up consistency checks within the survey so that inconsistent responses can be clarified immediately.
- Making questions measure-specific in order to improve the respondent's ability to provide concrete answers and to recognize that different measures installed at a project may have had different motivations.
- To estimate partial free-ridership, including questions that capture partial efficiency improvement (accounting for savings above baseline but less than program eligible) where applicable for a measure.

- To estimate partial free-ridership, including questions that capture program effects on the quantity purchased (accounting for situations where the customer would have installed some of the efficient equipment without the program, but not as much) where applicable for a measure.

- To estimate partial free-ridership, explicitly asking how the program affected the timing of the measure adoption, (accounting for situations where the measure would have been adopted without the program, but not as soon) where applicable for a measure.

Many of these design considerations also apply to the other kinds of survey efforts discussed in Chapter 2 that can be used to estimate free-ridership (or spillover) surveys.

b. Data collection elements

Even if stellar survey design elements are present, best practice data collection is key to collecting reliable and valid estimates. Key data collection elements include:

- Pre-testing the survey instrument to ensure questions are understandable, skip patterns are correct, and that the interview flows smoothly.

- Using techniques to minimize nonresponse bias, such as advance letters on utility letterhead and multiple follow ups over a number of weeks.

- Following professional standards for conducting surveys, and training and monitoring telephone interviewers (or engineers).

- Timing of the data collection should occur as soon as possible after installation to minimize recall bias and provide timely feedback on program design. Recognize, however, that timely data collection to estimate free-ridership will underestimate participant spillover and will increase data collection costs since a separate spillover survey would need to be conducted at a later date.

- Sampling a census (or oversample) of the largest savers and measures with few installations to ensure these are sufficiently represented in the survey sample.

c. Analytic elements

In addition to survey design elements, much of the literature discussed best practice analytic elements.

- Treatment of acceleration[16] to produce lifetime net savings rather than first-year net savings.

- Incorporating the influence of previous participation in the program. This incorporation recognizes that past programs participation may have had a positive impact on customers' behaviors and decisions to install additional equipment through the program.

- Establishing a priori rules for treatment of missing/don't knows in the scoring algorithm.

[16] Acceleration is discussed in more detail in Section 3.1.4 below.

- Weighting the estimates by annual savings to account for the size of the savings impacts for each customer.
- Calculating and reporting the precision of the estimate at the measure level.
- Algorithm characteristics and sensitivity testing of the scoring algorithm. Some of the methodologies reviewed relied on using different sets of questions to develop multiple scores of free-ridership, which are then averaged (or added together).
- Defining what the spillover measurement is and is not attempting to estimate, and why that approach was used.
- Employing, where feasible, a preponderance of evidence (or triangulation of results) approach that uses data from multiple sources, especially for large savers and complex decision-making cases. Data sources could include project file reviews, program staff and account manager interviews, vendor interviews, and site visit observations.

4.1.3 2003 standardized methodology

As noted in the 2003 standardized methodology report, there are fundamental differences between a penetration modeling approach over multiple years (using in, with, and without program scenarios) and an assessment of annual program impacts, including disaggregated values for free-ridership and spillover. The 2003 standardization project was designed to provide a methodology suitable for PAs to meet the regulatory requirements to report first year annual program impacts (along with disaggregated free-ridership and spillover values). This methodology provided an annual "snapshot" of the market as it currently operated.

However, the 2003 report noted that it is important to recognize that an annual "snapshot" of free-ridership and spillover measured in this way, without adequately considering the market effects associated with over a decade and a half of energy efficiency programs in Massachusetts, will result in net savings estimates that do not capture the full cumulative effect of the programs. As a result of these market effects, energy efficient technologies having high market share and few alternatives will have high free-ridership. That is, once the program has moved the market, continued program activity may have low attribution. At the same time, analysis of the current program alone does not necessarily credit the past program activity for its effect on the current market. That report noted that further work in documenting market effects was needed by the Sponsors before it could be considered along with the disaggregated free-ridership and spillover values.[17]

In the 2003 standardized report, the survey instruments and analysis were designed to estimate:

- Free-ridership (both pure and partial), using a customer survey

[17] At that time, the report noted that the Sponsors were undertaking other efforts to help document these C&I market effects, including a market share scoping study for HVAC and motors. The results of this study are contained in "Packaged Commercial HVAC Equipment Market Characterization, Final Report", June 30, 2006, KEMA, prepared for Northeast Energy Efficiency Partnership HVAC Sponsors.

- Participant like-measure spillover, using a customer survey (conservative approach)[18]
- Nonparticipant like-measure spillover, using a survey of participating design professionals and vendors (conservative approach).

In addition, the customer survey and analysis were designed to recognize the positive influences of program participation in prior years. Where past program influence was influential, the free-ridership estimate was reduced.[19]

4.1.4 Recommended changes to the 2003 standardized SRA methodology

While the 2003 standardized methodology follows many of the best practice question design, data collection and analysis elements discussed above in 3.1.2, we believe that some revisions can be made to the approach to improve the reliability and validity of the estimates. These changes were tested in the pretest discussed in Chapter 4 of this document.

1. The first recommendation deals with how acceleration is currently treated in Massachusetts. Unlike evaluation methods used in Wisconsin, California and New York, the Massachusetts method of treating acceleration differs from many acceleration treatments in that it gives full attribution credit for measures accelerated by more than one year, and no attribution credit for measures accelerated by less than one year. We suggest exploring acceleration in more detail by calculating life cycle net savings which determines the amount of savings attributable to the program over its lifetime. The procedure is as follows:

    a. For the reported acceleration period, calculate annual net savings as the difference in annual usage between the efficient technology and the old technology in place. Calculate total acceleration-period savings as the number of years of acceleration times this annual savings.

    b. For the time from the end of the acceleration period to the end of the measure life, calculate annual savings as the difference in usage between the efficient technology and the new technology that would have been in place without the program. If the same technology would have been bought without the program, this difference is zero. If standard efficiency technology would have been bought, annual net savings is the difference in annual usage between standard and efficient technology. Calculate total

[18] The 2003 standardized report noted that the measurement of spillover provided very conservative estimates of spillover, since one of the issues with quantification of spillover savings is how to value the savings of energy-efficient measures installed outside the program as a result of the program. Because this was an SRA approach, we had to rely on customer self-reports of the quantity and efficiency of any measures installed as well as vendor reports of like equipment sold outside the program. Resulting energy savings for like equipment could then be used from the program tracking system for that customer and that vendor to extrapolate spillover savings. This approach was suggested as experience has shown that while customers can usually report what type and quantity of equipment was installed and the operating hours, they typically cannot always provide sufficient information about the size and efficiency of that equipment to allow us to determine whether the equipment is "program-eligible" and what the savings should be. While a conservative estimate, this method does provide lower bound estimates that can be used for claiming participant spillover.

[19] Program participants were asked three questions about the influence past participation had on their decision to install the efficient equipment. If they agree to all three statements, the free-ridership rate is reduced by 75 percent. If they agree to two of the statements, the free-ridership rate is reduced by 37.5 percent.

post-acceleration-period savings as (measure life minus the number of years of acceleration) times this annual savings.

c. Calculate life-cycle net savings as the sum of the acceleration-period and post-acceleration period total net savings.

d. This method can be constrained by capping the acceleration period.

2. Another area to revisit is the timing of the interview. While the 2003 standardized methodology suggested completing the interviews as soon after the installation as possible, in practice, the interview was implemented anywhere from 3 to 15 months after installation.

   A Fast Feedback Pilot conducted for Energy Trust of Oregon[20] demonstrated that it is feasible to conduct free-ridership measurement on a rolling basis so that participants are contacted when the project is completed (or near completed). The advantages of this more frequent measurement include being able to provide more timely feedback to program managers, fewer issues with recall, and a higher likelihood that the decision maker will still be there. However, the authors note that there is an increased expense to this method.

   Based on our experience conducting rolling surveys, the added expense is not inconsequential, especially given the fact that there are multiple PAs. The additional expenses would be from: 1) additional costs and resource constraints for PA staff to provide monthly data in the appropriate format, 2) additional costs for the evaluator to select a sample from this monthly data[21], 3) additional interviewer training costs as not all interviewers will be available each month, 4) additional supervisory time as the surveys would not all be completed within one period of time, and 5) additional costs related to data analysis.

   More frequent data collection raises a couple other issues that need to be considered as well. For example, in cases where participants rate their vendor as being very influential in their decision to install efficient equipment, we would want to call the vendor back to discuss what impact they had on the decision. In annual studies, we are able to talk with influential vendors once about multiple projects where they were influential. If more frequent surveys were done, it would be necessary to call some vendors back multiple times during the year to discuss program influence. Not only vendors but also customers often have multiple projects. We prefer not to contact the same customer multiple times within a year. Contact too quickly after each project would mean that some potentially larger or more important projects would not be asked about because the customer had already been interviewed about a less important project.

   Additionally, data collection soon after the installation will further minimize observable spillover rates (see recommendation #7 below).

   For these reasons, we recommend interviewing respondents to estimate free-ridership no later than 6 months after installation. Ideally, we would recommend using quarterly sampling for prescriptive measures, especially large volume prescriptive, and semi-annual sampling for

---

[20] "Fast Feedback Pilot: Existing Buildings and Production Efficiency Programs, Final Report", prepared for the Energy Trust of Oregon by Research into action, March 10, 2010.

[21] The sampling costs can be mitigated to some extent by setting up routine processing systems. However, doing things on a more routine basis each round tends to result in less efficient sampling, in the sense that the same total data collection effort provides less statistically precise results.

custom measures (as custom measures are likely to have better recall and are more likely to be involved with large customers doing lots of things).

3. Some of the question batteries reviewed contained more extensive probing on other rival hypotheses for installing the efficient equipment. We recommend slightly revising the survey questions to incorporate probes to rule out alternative hypotheses.

4. A variety of different algorithms have been used to estimate free-ridership, including decision trees, additive and/or multiplicative rating scales, and multiple measurements of free-ridership which are averaged. The 2003 Standardized methodology algorithm produces a single estimate of free-ridership, based on responses to the timing, quantity, efficiency, consistency, technical assistance and past participation questions. As each different type of question is brought into the algorithm, the score is increased, decreased, or remains the same based on the response to the new question.

   More recent methodologies (e.g., the CPUC residential/small commercial and large commercial batteries) relied on using different sets of questions to develop multiple scores of free-ridership[22], from which an estimate is derived by averaging the different measurements. In theory, while this type of averaging improves the construct validity of the estimate, the use of an average scoring system means it is much less likely that a project will score as either 100% free-rider or 0% free-rider.

   We agree with some of the criticisms of using multiplicative approaches. By multiplying together two or more probabilities for several questions, each of which could be interpreted by itself as a measure of free-ridership, you only reduce the free-ridership rate. That being said, it is appropriate to use multiplication when adjusting for quantity installed.

   We recommend in the testing to use one primary set of questions to produce a free-ridership score, while using consistency questions (as well as other information where relevant for larger projects) to adjust the score.

5. In the 2003 algorithm the "don't know" responses to the initial free-ridership question are assigned a free-ridership value of zero percent. For these cases, we checked their responses to the consistency questions and their response to an open-ended question asking them to explain the program's impact in their own words, and then adjusted the free-ridership rate as appropriate. Other methodologies which rely on using different sets of questions to develop multiple scores of free-ridership (as discussed in point 4 above) drop the question set that contains don't knows. We will test these different methods as part of the algorithm testing discussed in Chapter 4.

6. Employing, where feasible, a preponderance of evidence (or triangulation of results) approach that uses data from multiple sources, especially for large savers and complex decision-making cases. In the 2003 methodology, we only talk with vendors if the customer says they were highly influential in the decision process. We recommend expanding this to talk with all vendors for the large savers. Using a case study approach to large and custom projects, we would also conduct in-depth interviews of key program staff and account managers who work regularly with a particular customer, in addition to a review of project files and notes. Further, we would include multiple in-depth interviews of key decision-makers where design teams are employed that may include A&E firms, contractors, customer staff from various functions, and others. Finally, any on-

[22] For example, see "Methodological Framework for Using the Self-Report Approach to Estimating Net-to-Gross Ratios for Nonresidential Customers" in Appendix B.

site observations from data collection contractors for M&V and impact evaluation may also be included for appropriate custom and otherwise large saver projects.

7. Changes to the existing method for estimating spillover (only 'like' spillover) will be more difficult, and expensive, to implement. As mentioned earlier, the current approach is a conservative method as it only asks about like spillover, and it only gives participants 3-15 months to install this additional equipment on their own. Asking participants about non-like spillover attributable to the program requires that we have sufficient information on the efficiency of the equipment to determine how much more efficient than the baseline it is.

   For example, in a past study done for National Grid, we did attempt to collect detailed information on equipment specifications from program participants, which we then forwarded to National Grid engineers to examine to determine the efficiency of the non-like equipment. While many respondents were able to provide at least some information, the engineers deemed that the information was insufficient for determining program eligibility and/or savings associated with the installation of the equipment. The only way to get this information would be to make on-site visits to inspect the equipment, which would be expensive unless these were done in conjunction with on-sites planned as part of the impact evaluation.

   A different approach would be to take a market share approach and talk with distributors, manufacturers, and retailers, as well as examine available market sales data. This could also be used to triangulate the SRA free-ridership estimates. This approach is effective only for large volume measures. As indicated in Section 2, a supply side study of market effects is currently underway for high bay lighting. The team will review with the PAs whether there are other measures that are large enough (in numbers of participants and overall savings levels) to justify the approach.

   Gathering information on like spillover from participants and participating vendors should continue to be done. Because the free-ridership measurement should be done as soon as possible after implementation, this does not leave participants with enough time to consider and install additional efficient equipment on their own. Therefore, we recommend splitting up the spillover and free-ridership measurement and following up on spillover one to two years after participation. An advantage to splitting these up is that it reduces the questionnaire length and respondent burden. The disadvantage of splitting up these measurements is a possible increase in participant burden (as many of these participants would have to be contacted twice), increased data collection and analysis costs, and increased chance that the original decision maker will no longer be with the company.

5. DEVELOP AND TEST PROPOSED METHODOLOGY

This section presents the results of tests that the Tetra Tech team used on the proposed revised free-ridership methodology. These tests were conducted in December 2010 and January 2011. These tests included pretest interviews with program participants and the use of simulated datasets.

First, this section summarizes in more detail the 2003 standardized free-ridership methodology that was reviewed as part of this methodology study, as well as the proposed changes to the 2003 methodology. This review is followed by the results of the pretest surveys conducted to test the revised instrument. As part of this pretest, the team completed 82 interviews with 2010 participants in National Grid's Small Business Services and Energy Initiative programs.

We then present the results of the simulated dataset test to estimate free-ridership. This test used two analysts to create simulated responses to the free-ridership questions related to hypothetical energy efficiency projects. Ten scenarios were created and scored. Out of those ten, the expected free-ridership estimates and the calculated free-ridership estimates matched for eight of the ten scenarios. Analysis of the mismatched scores indicates that the discrepancies are not the result of an inherent bias in the instrument or the scoring. We also present the spillover methodology and results from the pretest. This is followed in Chapter 5 by our recommended methodology to be used in the 2011 electric and gas studies with 2010 program participants.

5.1 FREE-RIDERSHIP METHODOLOGY PRETEST

The analysis and testing presented in this section focus on participant and contractor self-report methodologies for assessing free-ridership. As discussed earlier in this report, these two methods are not the only methodologies available for assessing free-ridership. However, as a majority of a Massachusetts programs will likely be assessed with participant self-report, the team prioritized the testing of those methodologies.

To assess free-ridership, the team proposed a survey instrument and scoring battery based on both the team's experience in the energy efficiency evaluation field and an extensive literature review of best practices in free-ridership from across the country. The proposed pretest instrument and scoring battery (referred to hereafter as the "2010 methodology") used the standardized methodology created for the Massachusetts programs in 2003 as a foundation (referred to hereafter as the "2003 methodology").

The team chose the 2003 methodology as a starting point because of its historical use in Massachusetts and because many of its approaches have been generally agreed upon by the PAs. In addition, the literature review conducted by the team in the fall of 2010 found that the 2003 standardized methodology followed many of the best practices in terms of question design, data collection, and analysis. Our review did identify a few revisions that could be made to the 2003 methodology to improve the reliability and validity of the estimates. These revisions include estimating free-ridership acceleration, probing to rule out rival hypotheses, the use of more consistency checks, missing data treatments, and "unlike" spillover. Our revised 2010 methodology addresses these concerns.

5.1.1 Summary of the 2003 Standardized Methodology

In order to understand how the proposed 2010 methodology differs from the previous 2003 methodology[23], we first briefly summarize the 2003 methodology. The 2003 methodology based its estimation of free-ridership and spillover on self-reported responses collected via a structured telephone survey. The survey instrument contained several distinct sections, all of which were used to measure and verify the participants' free-ridership at the measure category level. Table 5-1 describes each section of the instrument, its purpose, and how it is used in the final free-ridership estimate.

Table 5-1. Overview of 2003 Free-ridership Survey Instrument

Section	Purpose	Effect on Free-ridership
Identification of key decision maker(s)	Designed to ensure that the person(s) interviewed is the person who was most involved in the initial design and specification approval.	Only the correct respondent (or respondents) will be able to accurately speak about the decision.
Project and decision-making review	Used to assist customers in recalling the decision-making process as well as the services received through the program. Used along with the responses to the other consistency questions.	No direct effect but "warms-up" respondent and frames decision-making context and helps clarify free-ridership estimates in cases of inconsistent responses.
Initial free-ridership questions	Reminds the customer of the financial assistance they received for the measure. Customers are then asked about the impact of the Sponsor's incentive and technical assistance/education on the timing, quantity, and efficiency level of the specific measure category installed, as well as the amount of money they would have spent on their own in the absence of the program.	Classifies participants as pure free-riders, partial free-riders, or non-free-riders. See detailed discussion below. Initial classification is subject to adjustment by consistency checks and the influence of the technical assessment and previous program influence.
Consistency check questions	To improve the reliability of the total free-ridership estimates, these questions serve as consistency checks for those measure categories that are assigned an initial free-ridership rate of either 0% *or* 100%.	During the interview, every effort is made to resolve inconsistent responses. However, if during analysis, the initial free-ridership rate is clearly contradicted by the customer's response to the consistency question, the initial free-ridership is adjusted to 50% (from either 0% or 100%).
Influence of technical assessment (if applicable)	Assess the impact the utility's Technical Assessment Study had on participants.	For those customers who said they would not have paid the costs to have a similar study done on their own, their free-ridership rate is reduced by an adjustment factor of 50%.

[23] Pamela Rathbun, Carol Sabo, and Bryan Zent, PA Consulting Group. *Standardized Methods for Free-Ridership and Spillover Evaluation—Task 5 Final Report (Revised).* June 16, 2003

Section	Purpose	Effect on Free-ridership
Influence of past program participation	Recognize that past program participation in one of the Sponsor's energy efficiency programs may have had a positive impact on a customer's behaviors as well as their decision to install equipment through the program again. This adjustment is made to account for the market effects associated with over 20 years of energy efficiency programs in Massachusetts. These market effects will result in net savings estimates that do not capture the full cumulative effect of the program.	Depending on the degree to which previous participation has affected the participant, free-ridership is reduced either 75%, 37.5%, or not at all.
Participant "like" spillover questions	Ask about recent purchases (since program participation) of any additional energy-efficient equipment *of the same type* as that installed through the program due to program influences.	A spillover estimate is computed based on how much more of the same energy-efficient equipment the participant installed outside the program that was program influenced.

The initial free-ridership rate includes both pure and partial free-ridership. Based on the initial free-ridership questions, a measure category is classified as a pure free rider measure (100% free rider) if the customer indicates:

- They would have purchased the measure within 1 year (TIMING); AND
- They would have purchased the exact same amount (QUANTITY); AND
- All of the equipment they would have purchased would have been of similar efficiency level to what they installed through the program (PROGRAM EFFICIENCY); AND
- They would have paid for the entire measure cost (COST).

If one or more of the above four conditions are not met, the measure is assigned a non-free rider (0%) rate or a partial free-ridership (1-99%) rate. For measures where the customer indicates they would not have purchased the equipment within one year (TIMING), the free-ridership rate is 0%. For measures where the customer would have purchased it within one year but the quantity or efficiency would have been different or the customer would not have paid the full cost of the project; the partial free-ridership is calculated by applying the following formula to the customer's quantity and efficiency:

Initial FR =% QUANTITY * {% PROGRAM EFFICIENCY} where:

% QUANTITY = the percent of the equipment the customer would have purchased on their own without the program; and

% PROGRAM EFFICIENCY = the percent of the equipment the customer would have purchased that would have been of an efficiency level similar to (or higher than) the program equipment.

Initial classification is subject to adjustment by consistency checks and the influence of the technical assessment and previous program influence. After those adjustments, estimates are weighted to

represent population-level savings. These estimates can then be reported at the end-use and program level with corresponding confidence intervals.

5.1.2 Summary of the 2010 Proposed Methodology

While the 2003 standardized methodology follows many of the best practice question design, data collection, and analysis elements, the evaluation team identified several revisions that would improve the reliability and validity of its estimates. Table 5-2 lists each section from the 2003 survey instrument, the revisions that were made to the pretest of the 2010 methodology, and the rationale for that revision.

Table 5-2. Comparison of 2003 and 2010 Free-ridership Survey Instruments

Section	Revision	Reason
Identification of key decision makers	Minor revisions to question order and wording.	Reflects evaluation team's experiences with other data collection efforts In which the key decision-maker is identified more efficiently.
Project and decision-making review	Questions added regarding corporate policies, motivations for general equipment purchases, and the motivation for the incented project.	Used in consistency checks and as an additional test of the revised methodology.
Initial free-ridership questions	Additional questions about the project's budget, the program effect on the timing of the project, and the project's efficiency level. Restructured timing question to ask about installation at the same time instead of within 12 months.	Questions regarding the project's budget were added to aid in ruling out the rival hypothesis (i.e., the participant would have installed the same equipment at the same time without the program). Questions regarding the program's impact on timing were added to address the acceleration of savings. This addition allows the battery to more accurately capture the life cycle of savings from a particular measure. Questions regarding efficiency were added to clarify any partial free-ridership.
Consistency check questions	In addition to an open-ended consistency check question, scale questions about the overall likelihood of the project without the program, the influence of the incentive, and the influence of any third-party contractors were added.	These questions were added to aid in ruling out the rival hypothesis (i.e., the participant would have installed the same equipment at the same time without the program) when inconsistent responses are given.
Influence of technical assessment	Added a scale question about the influence of the technical assessment.	Used instead of binary Yes/No item and reflects the hypothetical nature of the question.
Influence of past program participation	No changes.	
Participant "like" spillover questions	Minor changes to "like" spillover and addition of "unlike" spillover questions. Recommend adding a 0-10 influence scale item.	Added to assess the amount of "unlike" spillover that may be occurring due to program influence. Scale item added as consistency check.

Based on changes made in the 2010 methodology, the team expected higher rates of free-ridership than were found in the 2003 methodology, as the 2003 methodology did not adjust for acceleration. As discussed above, any project assessed using the 2003 methodology that would not have been completed within a year of participation was scored as a non-free-rider. In contrast, the 2010 methodology increases that limit to 24 months for small business programs and 48 months for large commercial programs.[24] For example, using the 2003 methodology, participants that reported that they would not have installed any equipment within 12 months would not be considered a free-rider. However, in 2010, the same participants would be considered *partial* free-riders. Other changes to the 2010 methodology, such as additional consistency checks, the scale items, and minor wording edits may affect the estimates but not in a consistent direction.

5.1.3 2010 Methodology Pretest Survey Questions

The pretested 2010 instrument begins as the 2003 instrument did. The interviewer first identifies the appropriate decision-maker within the organization by asking if participants were involved in the decision to purchase the rebated measure and the roles of others in the organization that may have been involved. The question text used in the pretest is detailed below.

I1 Are you the person who was most involved in making the decision to get <ALL PROGRAM ASSISTANCE> through the <PROGRAM> in <DATE> at <ADDRESS> in <CITY>?

I1A Who was primarily responsible for making the decision to get <ALL PROGRAM ASSISTANCE> through the program?

I2 Do you work directly for <CUSTOMER> or are you a contractor who provides design and/or installation services for <CUSTOMER>?

The interview then asks about corporate purchasing policies, important factors that he or she considers when purchasing any new equipment, and important factors for the specific rebated project. As in the 2003 methodology, this section is intended to "prime" the participant by asking them to recall factors that may have been important in the purchase decision. The question text is listed below.

R3 Does your company have any corporate policies related to energy efficiency standards that you need to consider when purchasing new equipment or making improvements to this facility?

R4 Which of the following best describes this policy: purchase energy efficient measures regardless of cost, purchase energy efficient measures if it meets payback or return on investment criteria, purchase standard efficiency measures that meet code, or something else?

R4B When purchasing new equipment for your facility such as <MEASURE CATEGORY 1> or <MEASURE CATEGORY 2>, what is the most important factor in your decision on what type of equipment to purchase?

R4C What is the second most important factor in your decision?

[24] The California methodology uses 4 years for large C&I customers and 2 years for small C&I customers to more accurately reflect the length of time involved in planning different sizes of projects.

R5 Please think back to the time when you were considering implementing the <MEASURE CATEGORY 1 and MEASURE CATEGORY 2> projects. What factors motivated your business to consider implementing new <MEASURE CATEGORY 1 and MEASURE CATEGORY 2> equipment?

The instrument then discusses what influence, if any, the program had on the decision to install equipment through the program. As there are several dimensions to the decision to purchase and install new equipment[25], the battery discusses the timing of the installation and the quantity and the efficiency level of the equipment installed. These questions include both the overall effect of the program (including staff recommendations and any technical assistance) and the specific effect of the financial incentive. The questions are listed below. Please note that these questions are measure-specific and are repeated for up to two measures.

FR5 According to our records, the total cost for the project implemented at your facility in <DATE> through the <PROGRAM> was about <TOTAL PROJECT COST>. <PROGRAM ADMINISTRATOR> paid about <INCENTIVE> of the total cost of the [IF EFFECIENCY APPLIES: ENERGY EFFICIENT] <MEASURE CATEGORY> project implemented through the program.

[IF NO <STUDY>: You may have also received some technical assistance from a <PROGRAM ADMINISTRATOR> rep, engineer, or equipment vendor.]

If <PROGRAM ADMINISTRATOR> had not paid a portion of the implementation cost OR provided any technical assistance or education through the <PROGRAM>, would your business have implemented any type of <MEASURE CATEGORY> project at the same time?

FR6A Would you have implemented the <MEASURE CATEGORY> project earlier than you did, at a later date, or never?

FR6B How much [EARLIER/LATER] would you have implemented the <MEASURE CATEGORY> project?

FR7A Without the program incentive and technical assistance or education, would your business have implemented the exact same quantity of <MEASURE CATEGORY> equipment [IF FR5=YES OR DK: AT THAT SAME TIME; IF FR5=2: WITHIN (TIMEFRAME IN FR6B)]?

FR7B Compared to the amount of <MEASURE CATEGORY> that you implemented through the program, what percent of the project do you think your business would have purchased on its own during that timeframe?

FR8A You said your business would have installed [IF FR7A=YES: all; IF FR7A= NO: (FILL WITH FR7B %)] of the equipment on its own if the program had not been available.

Thinking about the <MEASURE CATEGORY> equipment you would have installed on your own, what percent of this equipment would have been of the same high efficiency as what was installed through the program?

[25] The instrument is designed to handle both rebated equipment (e.g., lighting) and rebated services (e.g. delamping). However, as the pretest only addresses equipment, the memo does not include any references to rebated services.

FR8B Of lower efficiency than what was purchased but higher than standard efficiency or code?

FR8C[26] And of standard efficiency or code?

The pretest also included questions that would confirm or correct inconsistent responses. For example, if participants reported that they were likely to install the equipment without the program but also reported that they would not have implemented the equipment within four years, the interviewer asked them to confirm which statement was more accurate. These questions are listed below.

FR1 On a scale of 0 to 10, with 0 being not at all likely and 10 being very likely, how likely is it that your business would have implemented the same [IF QUANTITY VARIES: QUANTITY AND] [IF EFFICIENCY APPLIES: EFFICIENCY OF] <MEASURE CATEGORY> at that same time if the <PROGRAM ADMINISTRATOR> had not provided the <PROGRAM ASSISTANCE>?

C3 On a scale of 0 to 10, with 0 being no influence and 10 being a great deal of influence, how much influence did the <INC> you received from <PROGRAM ADMINISTRATOR> have on your decision to implement the [IF EFFICIENCY APPLIES: HIGH EFFICIENCY] <MEASURE CATEGORY> project?

C4A Now I want to focus on what it would have cost your business to install this equipment on its own without the program. On a scale of 0 to 10, with 0 being not at all likely and 10 being very likely, how likely is it that your business would have paid the additional <INC> on top of the amount you already paid, to implement the same quantity and efficiency of <MEASURE CATEGORY> equipment at that same time?

C8 [ASK IF FR1 > 3 AND FR6b >24/48 MONTHS OR NEVER] Earlier in the interview, you said there was a [FR1 SCORE] in 10 likelihood that you would have implemented the same quantity and efficiency of <MEASURE CATEGORY> equipment at that same time in the absence of the program assistance. But you also said you would not have implemented the <MEASURE CATEGORY> project within 4 years of when you did. Which of these is more accurate?

C9 I'd like to better understand your purchase decision. In your own words, please describe what impact, if any, the program had on your decision to install the energy efficient <MEASURE CATEGORY> equipment at the time you did and in the quantity you did?

5.1.4 2010 Methodology Pretest Scoring

Prior to fielding, the team constructed a scoring system based on the influence and consistency check questions above. The scoring calculates two scores: a quantity score and an efficiency score. The quantity score represents the percentage of the rebated equipment that would have been installed in absence of the program. The efficiency score is the percentage of savings *per unit installed* that would have occurred without the program. For equipment that is reported to be more efficient than standard but less efficient than what was installed through the program, we assume 50 percent of the savings for those measures Multiplying these two scores together gives the percent of the rebated savings that would have occurred without the program. This percentage is the raw free-ridership estimate. Table 5-3 details these calculations.

[26] For measures where quantity is not applicable but efficiency levels do vary, this question is combined into one item: FR8D.

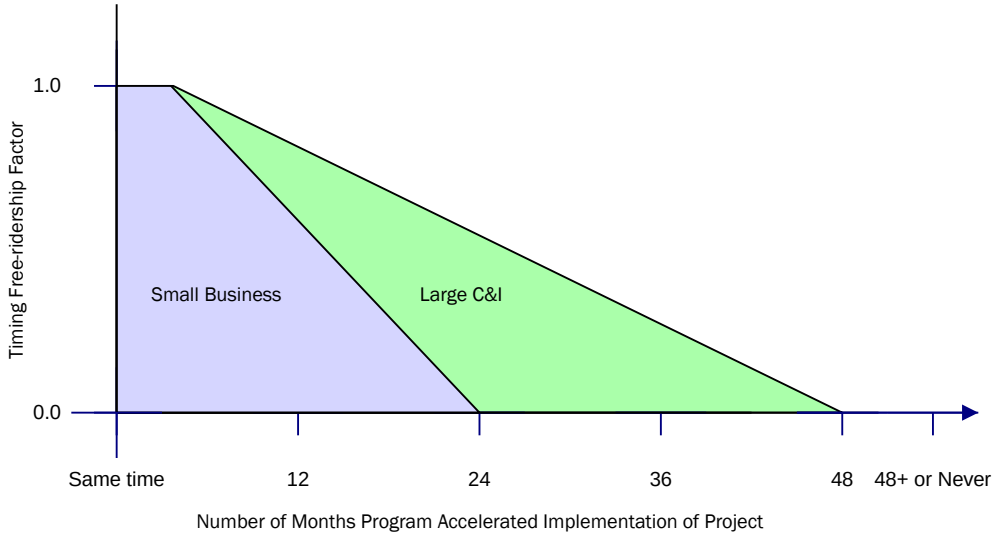
Table 5-3. Quantity and Efficiency Scores

Score	Responses	Result
Quantity Score (FR_QTY)	If would have installed same quantity without program (FR7A = YES)	FR_QTY = 1
	If would have installed fewer quantity without program (FR7A = NO)	FR_QTY = FR7B
	If never would have installed (FR6A = never)	FR_QTY = 0
Efficiency Score (FR_EFF)	If would have installed at least some equipment on their own	FR_EFF = FR8A + (FR8B*.50)
	If never would have installed (FR6A = never)	FR_EFF = 0
Initial Free-ridership Score	The percent of the rebated savings that would have occurred without the program.	FR_EFF * FR_QTY

The product of these two scores is then adjusted by a timing factor. The timing factor adjusts the raw free-ridership estimate downward for all or part of the savings that would have occurred without the program, but not until much later. That is, the program is given credit for accelerating the installation of energy efficient equipment. For example, if the participant states that he or she would have installed equipment at the same time regardless of the program, the quantity-efficiency factor is not adjusted. However, if the participant states that, without the program, they would have completed the project more than 6 months later than they actually did, any free-ridership identified in quantity-efficiency factor is adjusted downward[27]. The degree of the adjustment depends on the program. As the equipment planning schedule for small businesses is likely shorter than the planning schedule for large business, small business programs receive a greater acceleration benefit. This reduced adjustment for small businesses reflects the increased effect the program has on the planning schedule. This adjustment is detailed in Table 5-4 and visualized in Figure 5-1.

[27] Projects that were accelerated by fewer than 6 months are not adjusted. As installation timelines are subject to shifting, we assume these projects are just as likely to have been installed at the same time.

Figure 5-1. Timing Free-ridership Factor by Number of Months the Program Accelerated Implementation



This adjusted score is reviewed for consistency and for vendor influence via a follow-up interview with vendors that are rated influential by participants.

Table 5-4. Timing Factor Adjustment

Score	Responses	*Result*
Timing Factor – Small Business Programs (FR_TIMING)	Would have installed at the same time without the program (FR5 = Yes)	*FR_TIMING = 1*
	Would have installed within six months of when participant actually did without the program (FR6A <= 6 months)	*FR_TIMING = 1*
	Would have installed sometime between 7 and 24 months of when participant actually did without the program (FR6A > 6 months & < 24 months)	*FR_TIMING = 1-((FR6B-6) * .056)*
	Would have installed sometime after 24 months of when participant actually did without the program (FR6A > 24 months)	*FR_TIMING = 0*
	Would have never installed without the program (FR6A = Never)	*FR_TIMING = 0*
Timing Factor – Large Business Programs (FR_TIMING)	Would have installed at the same time without the program (FR5 = Yes)	*FR_TIMING = 1*
	Would have installed within six months of when participant actually did without the program (FR6A ≤ 6 months)	*FR_TIMING = 1*

Score	Responses	*Result*
	Would have installed sometime between 7 and 48 months of when participant actually did without the program (FR6A > 6 months & < 48 months)	*FR_TIMING = 1-((FR6B-6 * .024)*
	Would have installed sometime after 48 months of when participant actually did without the program (FR6A > 48 months)	*FR_TIMING = 0*
	Would have never installed without the program (FR6A = Never)	*FR_TIMING = 0*
Adjusted Free-ridership Score	*The raw free-ridership estimate adjusted for all or part of the savings that would have occurred without the program, but not until much later*	*FR_TIMING * Initial Free-ridership Score*

Finally, the score is further adjusted by the influence of any program-sponsored technical assistance or audit and by the influence of previous program participation. If a participant rates the influence of the technical assistance as high (7 or greater on a scale of 0-10), the free-ridership score is reduced by half. This reduction is necessary because the previous factors focus on the specific effect of the program incentive and the overall effect of the program. Without this adjustment, the influence of the technical assessment is under-represented.

C2 On a scale of 0 to 10, with 0 being no influence and 10 being a great deal of influence, how much influence did the information provided by the <STUDY> have on your decision to implement the [IF EFFICIENCY IS APPLICABLE*:* high efficiency] <MEASURE CATEGORY> project?

Likewise, if a participant has previously participated in the program, they are asked about the influence of that past participation on their perceptions and behaviors. Participants are asked to state whether they agree or disagree with four statements about the effect past participation has had on their decisions. Based on the number of statements with which they agree, their free-ridership is reduced by 75 percent, 37.5 percent, or not reduced at all. This reduction is done to account for the influence positive program experiences have had on participants' purchasing decision – with the program administrators, implementers, or the equipment incented.

PP3. I'm going to read you several statements. For each statement, please tell me whether you agree or disagree that this statement applies to your business. There are no right or wrong answers; we just want your honest opinion.

Our previous experience implementing energy efficient projects through the <PROGRAM> . . . .

a. Has made our firm more likely to consider energy efficient equipment
b. Has made our firm more likely to install energy efficient equipment
c. Has given us more confidence in the financial benefits of energy efficient equipment
d. Has given us more confidence in the nonfinancial benefits of energy efficient equipment

As mentioned previously, the previous program participation adjustment is made to account for the market effects associated with over 20 years of energy efficiency programs in Massachusetts. These market effects will result in net savings estimates that do not capture the full cumulative effect of the program. The 2003 methodology attempted to capture some of these market effects by making this adjustment for previous program participation. While it could be argued that the influence of previous participation should count as spillover rather than reduced free-ridership, the traditional definition of spillover does not count measures installed through a program as spillover. Table 5-5 details these adjustments.

Table 5-5. Adjustments for the Influence of Technical Assessments and Previous Participation

Adjustment	Responses	Result
Technical Assessment Adjustment	No technical assessment, audit, or study conducted	No adjustment
	Participant would have performed assessment, audit, or study without program assistance or it was not influential (C2 ≤ 6)	No adjustment
	Participant **would not** have performed assessment, audit, or study without program assistance and it was influential (C2 > 6)	Adjusted Free-ridership Score * .5
Previous Participation Adjustment	No previous participation in program	No adjustment
	Agrees with four statements regarding the positive influence of past participation (PP3)	Adjusted Free-ridership Score * .25
	Agrees with three statements regarding the positive influence of past participation (PP3)	Adjusted Free-ridership Score * .625
	Agrees with two or fewer statements regarding the positive influence of past participation (PP3)	No adjustment

Flowchart diagrams detailing these calculations have been included in Appendix F of this report.

5.2 2010 METHODOLOGY PRETEST DATA COLLECTION

This section presents the sampling, methodology, and results of the revised free-ridership and spillover pretest developed by the Tetra Tech team. The revised instrument was fielded in December 2010 after a literature review of free-ridership evaluation techniques in the fall of 2010. As part of this pretest, the team completed 82 interviews with 2010 participants in National Grid's Small Business Services and Energy Initiative programs.

5.2.1 Sampling

The Tetra Tech team selected a random sample of National Grid's Energy Initiative and Small Business Services 2010 participants to be part of the revised instrument pretest. In the Energy Initiative program, participants that received rebates for lighting or compressed air projects were included in the sample. In the Small Business Services program, only lighting projects were included.

National Grid participants were selected for two reasons. First, National Grid conducted a free-ridership study of its 2009 participants in the spring of 2010. This recent study could be used as a point of comparison with the pretest results. In addition, the National Grid programs have a large pool of measures and participants from which to sample. Our inclusion of lighting participants in our pretest would not exhaust the sample for any subsequent evaluation. It is for this reason that the team did not pretest the instrument with Custom projects as originally planned. We did not want to exhaust the small participant population for the 2011 study.

The team drew the sample from two population files provided by National Grid on November 10 and November 24. Before randomly selecting the sample, the population was aggregated by program, measure, and account number. To ensure that later evaluation efforts had a sufficient pool of records from which to sample, no priority accounts (i.e., "high savers") were selected.

Calling occurred between December 9 and December 23, 2010. Prior to calling, participants were sent an advance letter on National Grid letterhead explaining the purpose of the call and inviting them to contact Tetra Tech at their convenience on a toll-free number. On average, the interview lasted 16 minutes. Table 5-6 details the cooperation rate for the pretest study by program and overall. This rate is lower than what we would normally target (typically 65 percent). Given the small sample and short timeline, the team placed emphasis on completing surveys within a two week period.

Table 5-6. Cooperation Rate with the 2010 Pretest

Status	Energy Initiative	Small Business Services	Total
Original Sample	**88**	**87**	**175**
No working number	5	5	10
Vendor/Contractor	0	1	1
Adjusted Sample	**83**	**81**	**164**
Refusal	6	3	9
Active Sample	32	41	73
Completed Surveys	**45**	**37**	**82**
Cooperation Rate	**54%**	**46%**	**50%**

5.2.2 Results

Table 5-7 summarizes the free-ridership rates by program and measure using the scoring procedures described earlier in this document. As part of this pretest, the free-ridership rate identified for the lighting projects in National Grid's Energy Initiative program is 17.4 percent. For compressed air projects, the rate is slightly higher - 26.7 percent. For National Grid's Small Business Services program, the rate for lighting projects is 3.6 percent.

These rates are comparable with the rates identified in the full-scale 2009 study that used the 2003 methodology. However, there are two factors that make a comparison between the rates using the 2003 methodology and the 2010 rates difficult. First, the question wording in the 2010 instrument is different from that of the 2003 instrument. This change may lead to some fluctuations in the participants' responses. In addition, the lag between participation in the program and the interview is reduced. The 2009 evaluation using the 2003 methodology occurred in the spring of 2010, meaning that for some participants, there may have been a 16 month lag before the interview was conducted. In contrast, the pretest calls occurred within 11 months of participation. The evaluation team speculates that reducing the delay in evaluation may allow for the participants to provide more accurate responses and possibly have better recall as to what influence the program had on their decision to install the equipment. We should note, however, that this shortened time period will also reduce any identified spillover.

Table 5-7. 2010 Free-ridership Rates

Program	Measure	2010 Free-ridership Rate[28]	2009 Participant Free-ridership Rate using 2003 Methodology	Number of 2010 Participants
Energy Initiative	Lighting	17.4%	11.9%	39
	Compressed Air	26.7%	23.8%	6
Small Business Services	Lighting	3.6%	6.2%	36[29]

5.2.3 Different Scoring Scenarios

The rates in Table 5-7 above use median substitution to adjust for missing responses to the quantity and efficiency questions and include the deferred attribution adjustment that accounts for accelerated installations. The median is used because valid responses to the quantity and efficiency questions are not normally distributed; instead they are clumped around 0 and 1.0. We have included an analysis that substitutes the mean, rather than median values, for missing responses in Appendix D for comparison.

Table 5-8 below presents a comparison of the free-ridership rates when calculated without median substitution. In this scenario, if the participant was unable to answer questions regarding quantity and efficiency levels, the record was removed from analysis. This scenario results in the removal of seven Energy Initiative lighting projects records where the respondent was unable to estimate the efficiency level installed absent the program.

Free-ridership estimates for lighting that use median substitution are higher than estimates based on data that exclude records with missing responses. This difference is a result of which items participants are most likely to have difficultly recalling and our interpretation of their difficulty. In our methodology, we interpret responses of “don’t know” to the timing question as “never”. That is, if the participant cannot estimate an installation time frame, we assume they are unlikely to have installed the energy efficient equipment within the 2/4-year timeframe without the program. This assumption means these participants are coded as zero percent free-riders (subject to consistency checks).

The evaluation team believes that using median substitution is the more credible and the more conservative estimate of free-ridership. As median substitution is only used for participants that report they would have implemented at least some portion of the project without the program, we believe assuming that participants that provide a response of “don’t know” to either the quantity or efficiency items question have some level of free-ridership is appropriate. Please note that, during the interview, interviewers probe “don’t know” responses in order to get the participant to give his or her best estimate. Also, during analysis, cases that use impute median substitution values are subject to the same consistency checks used for other cases.

However, we do not make similar assumptions for the “don’t know” responses to the efficiency and quantity questions. If a participant is unable to estimate the amount of efficient equipment that he or she would have installed without the program, we do not believe it is a reasonable assumption to assume that he or she would not have installed any efficiency equipment absent the program (thus

[28] Free-ridership rates are weighted by gross kWh savings.
[29] One lighting project was removed from analysis because the equipment was no longer installed.

implying a zero percent free-rider). Therefore, using median substitution of responses for the efficiency and quantity questions effectively raises the free-ridership rate.

Table 5-8. Different Free-ridership Scoring Scenarios

		Free-ridership with median substitution	N	Free-ridership without median substitution	N
Energy Initiative	Lighting	17.4%	39	4.4%	32
Energy Initiative	Compressed Air	26.7%	6	26.7%	6
Small Business Services	Lighting	3.6%	36	3.6%	36

A second scoring scenario, presented in Table 5-9, attempts to replicate the 2003 methodology with the data collected as part of the pretest.[30] In essence, this analysis removes the acceleration adjustment – in the 2003 methodology, projects that would have been completed within 12 months are considered free-riders while those projects that would have been completed outside of 12 months of participation are not. Removing this adjustment has the expected effect of decreasing free-ridership as discussed previously.

Table 5-9. Replication of 2003 Methodology

		Free-ridership with median substitution	N	Replicated 2003 methodology	N
Energy Initiative	Lighting	17.4%	39	14.5%	39
	Compressed Air	26.7%	6	13.3%	6
Small Business Services	Lighting	3.6%	36	0.1%	36

5.2.4 Stability Analysis

As part of the development of the new instrument, the team tested the pretest data for both reliability and validity. A reliable instrument would produce consistent results across applications. A valid instrument would measure the constructs that we intended to measure.

First, to test reliability within the free-ridership battery, the team calculated a Cronbach's alpha of .778 for the free-ridership score and two independent consistency checks: FR1 (the likelihood that the participant would have installed the same project without the program) and C3 (the influence of the incentive on the participant's decision to implement the project recoded to match the direction of the free-ridership rates). As Cronbach's alpha approaches 1.0, the intercorrelations between the test items increase. Therefore, this test can be interpreted as a measurement of the internal reliability of a set of

[30] Ideally, our test design would have randomly assigned participants to either receive the 2010 methodology or the 2003 methodology. However, due to sample constraints, this treatment was not possible.

items. In this situation, Cronbach's alpha suggests that FR1, C3, and the calculated free-ridership score are relatively reliable in estimating the level of free-ridership reported by a participant.

We also tested the validity of the instrument, that is, do the instrument and scoring provide accurate or "true" assessments of free-ridership levels? We performed three tests: a comparison to previous rates, a comparison of average free-ridership rates by participant characteristics, and correlations between participant characteristics and free-ridership rates.

Our first test was to compare the results to the previous evaluations as in Table 5-7 above. This test assumes that the results of the evaluation of 2009 participants using the 2003 methodology were valid. This test is useful as a comparison of how much change the revised methodology has introduced when compared to the 2003 methodology. As previously discussed, the rates found at the measure level are higher than those identified in 2009 suggesting that our instrument has adjusted free-ridership in the direction we expected.

We also examined the calculated free-ridership rate broken out by several items that were asked as warm-up questions to the free-ridership battery. These items were chosen because they were not included in the scoring and they are related to attributes and priorities that should affect the purchasing decision of energy efficiency equipment. These items include the existence and type of corporate policy regarding energy efficiency, the most important factor that the participant considers when purchasing new equipment, and what factors motivated the participant to implement the rebated project.

Significance tests were calculated for these differences of means when applicable. However, due to the small sample sizes, none of the differences are statistically significant. Therefore, these results are presented to construct a qualitative, "preponderance of evidence" approach to validity.

First, a comparison of average free-ridership rates for those participants that reported having a corporate policy regarding energy efficiency with those participants that do not have such a policy suggests our measurement is valid. As detailed in Table 5-10 and Table 5-11, participants that reported having a policy regarding energy efficiency had a higher level of free-ridership than those that did not. In addition, participants whose policy specified that they purchase energy efficient measures regardless of cost or achieve a percentage reduction in energy usage annually exhibit higher levels of free-ridership. Both of these results suggest that our measurement of free-ridership is valid. Please note that our analysis is limited by both low free-ridership rates and a small sample. Therefore, we are unable to examine customer characteristics with great depth or rigor. However, our analysis is generally indicative of a correct assessment of free-ridership.

Table 5-10. Free-ridership by Corporate Policy Regarding New Equipment

R3 - Company has energy efficiency policy when purchasing new equipment	Free-ridership Rate	Number of Participants
Yes	32%	19
No	10%	60
Don't know	0	2
Total	**16%**[31]	**81**

[31] Differences are not statistically significant.

Table 5-11. Free-ridership by Type of Corporate Policy

R4 - Type of energy efficiency policy at company	Free-ridership rate	Number of Participants
Purchase energy efficient measures regardless of cost	11%	1
Purchase energy efficient measures if it meets payback or return on investment criteria	7%	16
"All of the above"	30%	1
Corporate policy requires annual energy reduction	100%	1
Total	**32%**[32]	**19**

In addition, analysis of the free-ridership rates and the factors participants reported most important in their decision to purchase new equipment suggests that our scoring is reasonably assessing free-ridership. Participants who reported that "equipment performance" and "equipment energy consumption" were the most important factors in their purchase decision displayed the highest free-ridership rates. As both of these factors are generally unaffected by program incentives or assistance, participants that prioritize them above cost (the barrier programs frequently work to overcome) would likely purchase the same equipment without program encouragement. Likewise, participants that reported "upfront cost" and "payback" as the most important factors in their purchase decisions have comparatively low free-ridership rates. Again, as program influence directly affects these two factors, this pattern suggests that the instrument and scoring are correctly assessing free-ridership. Table 5-12 details the comparison of free-ridership rates by factor.

Table 5-12. Free-ridership by Important Factors in New Purchases

R4B - Most important factor in new equipment purchases	Free-ridership Rate	Number of Participants
Equipment performance	37%	11
The equipment's energy consumption	20%	26
Contractor, design professional, engineer's recommendations	10%	2
Lifetime cost of the equipment	12%	4
First (or up-front) cost of the equipment	11%	17
Payback	7%	10
Something else	1%	5
Purchasing 'green' equipment	0%	1
Program administrator recommendations	0%	1
Meeting code	0%	1
Don't know	0%	3
Total	**16%**	**81**

[32] Differences are not statistically significant.

Also, when we compared the average free-ridership rate by which motivational factors the participant reported in his or her decision to implement the rebated project, those participants that did not report any program-influenced factors had, on average, higher free-ridership rates. Because participants could report more than one factor, we coded them into two groups: those that reported **at least one** program-influenced factor and those that **only** reported factors outside the program's influence. We considered the following factors within the program's influence: the program incentive, any technical assistance offered by the program, recommendations by third party contractors (the programs work closely with trade allies), recommendations by program administrator staff, and past program experience. Factors outside of the program's influence were equipment failure, equipment performance, a reduction in maintenance costs, a reduction in energy usage, and a reduction in energy costs.

When grouped in this manner, participants who did not mention any program-influenced factor had a higher free-ridership score than those that did. Again, as those participants who are motivated by factors outside of the program's influence are likely to have installed a similar project in absence of the program, this pattern suggests that our instrument and scoring are assessing free-ridership in the direction we would expect. Table 5-13 summarizes this comparison by group.

Table 5-13. Free-ridership by Factors in Implementation Decision

R5 - Factors in decision to implement rebated projects	Free-ridership Rate	Number of Participants
Factors in decision included program influence	9%	31
Factors in decision outside of program influence	18%	50
Total	**16%**[33]	**81**

Finally, correlations among participant characteristics and the free-ridership rates, while not strong, are in a direction that suggests our scoring to be reasonably assessing free-ridership. By creating yes/no "dummy" variables based on participants' responses to several questions regarding the purchase decision, we were able to test the correlation with our identified free-ridership rate. As mentioned earlier, our sample size is too small to achieve statistical significance in this test. However, we are able to test whether the direction of the association is in the direction we would expect given the characteristics relationship with free-ridership. Table 5-14 summarizes the correlation coefficients for each characteristic.

Table 5-14. Free-ridership Correlations by Participant Characteristics

Participant Characteristics	Pearson Correlation with Free-ridership Rate (n = 81)
Cost primary concern when making new purchases	-0.123
Equipment primary concern when making new purchases	0.193
Only equipment factored in decision to implement	0.047
Corporate policy on energy efficiency	0.000
Pre-existing budget before speaking with the program	0.325[34]

[33] Differences are not statistically significant.
[34] Only the "pre-existing budget" item shows a statistically significant correlation to free-ridership.

For example, we expect that reporting cost as a primary concern when purchasing new equipment would indicate low free-ridership as the program's relatively large incentives directly mitigates a project's upfront cost. A negative correlation between free-ridership and this characteristic suggest our scoring is working in the right direction. Conversely, the existence of a budget before the participant talked with the program would indicate some level of free-ridership as the participant may have participated in the program after already committing to implementing the project without knowledge of the incentive. A positive correlation between that characteristic and free-ridership again suggests that our instrument and scoring are correctly assessing free-ridership.

5.3 SIMULATED PRETEST DATASET ANALYSIS

The simulated dataset test we employed is similar to a recent approach used to assess reliability for self-reported net-to-gross surveys, which team members were involved with in the 2006-2008 California impact evaluations[35]. First, two analysts created a simulated dataset of purchase scenarios. One analyst specified the scenarios, a set of motivations, and a subjective range of free-ridership estimates associated with each scenario. The second analyst was given the scenarios (scrubbed of the range of free-ridership estimates and independently responded to the survey questions for each scenario. Finally, the first analyst applied the specified net-to-gross calculation algorithm to the survey responses and compared the results to the pre-established subjective free-ridership estimates. The results of this testing, presented in Table 5-15 below, were used to further test the validity of the survey instrument and the analysis algorithm.

[35] Lori Megdal, Megdal & Associates, LLC; Pam Rathbun, Laura Schauer, Jeremy Kraft, and Kimberly Bakalars, PA Consulting Group, Inc. Testing Residential/Small Commercial Free-ridership Algorithm for the JS NTG instrument & SAS Coding Review and Revisions. Memo to the CPUC. July 27, 2009.

Table 5-15. Results of Simulated Analysis

Case	Scenario	Scored Free-ridership Estimate Based on Simulated Responses to Hypothetical Scenarios	Subjective Free-Ridership Estimate Based on Hypothetical Scenarios	Match?
1	HVAC system failed, contractor sold customer on upgrade to high efficiency with incentive. Focused on first cost.	0%	0%	Yes
2	HVAC system failed, wanted to replace with high efficiency. Focused on efficiency and lifetime cost.	100%	100%	Yes
3	Small business - Purchased lighting for front and back offices of facility instead of just back due to rebate. Direct installed. Focused on efficiency and lifetime cost.	50%	50%	Yes
4	Small business - Leases facility, pay energy bills but have difficulty upgrading. Upgrade motors. Focused on first cost.	0%	0%	Yes
5	**Large C&I - corporate policies - replace on failure. Has worked with program in the past. Focused on efficiency.**	**100%**	**75%**	**No**
6	Not failed equipment – money in budget for upgrades, investigating lighting options. Used technical assistance and focused on first cost.	0%	0%	Yes
7	**Medium customer with aging HVAC equipment. Used technical assistance and focused on first cost.**	**0%**	**25%**	**No**
8	Large customers with aging motors. Focused on lifetime cost.	100%	100%	Yes
9	**Large customer with custom building-wide project. Limited funds. Used technical assistance and focused on first cost.**	**6%**	**0%**	**No**
10	Medium customer with working equipment. Looking to position them as green in the market. Focused on efficiency.	100%	100%	Yes

From our comparison, seven of the ten hypothetical scenarios exactly match the predicted free-ridership score. Three scores do not match – the cases bolded in the table above. The analysis team examined those cases after scoring to try and determine what in the scoring caused the scores to differ.

For Case 5, the researcher creating the simulated set of responses (referred to as "the response team") did not respond to questions regarding the effect of previous program experience. Therefore, the scoring did not reduce the free-ridership to reflect the influence of previous program experience.

For Case 7, the expected free-ridership estimate was based on the scenario that this participant was focused on both the first cost of the equipment and the lifetime cost. Given these motivations, the response team stated that the participant would have purchased standard efficiency equipment without the program. This response indicates that the participant is a zero percent free-rider. However, the researchers that created the expected free-ridership score assumed that the participant would have purchased high efficiency equipment, just at a later date. This assumption was based on the lifetime cost motivation.

For Case 9, the expected free-ridership estimate and the scored estimate only differ by six percentage points. Given the subjective nature of this analysis, we feel this difference is within the margin of error and does not suggest any flaw or bias.

Our simulated estimates exactly match a majority of the expected estimates. When they do not match, the differences do not show a bias in either direction – one estimate is high by 25 percentage points, another is low by 25 percentage points, and the third within the margin of error. Therefore, we feel that this analysis suggests that there is not systematic bias in the instrument's assessment of free-ridership.

5.4 SPILLOVER PRETEST METHODOLOGY

The 2003 methodology measured two types of spillover: participant "like" and nonparticipant spillover. Participant "like" spillover were the estimated savings that were attributable to the program by program participants that installed equipment outside of the program *of the exact same type* as what was installed through the program. Nonparticipant spillover is energy efficient measures installed by program nonparticipants due to the program's influence. The program can have an influence on design professionals and vendors as well as an influence on product availability, product acceptance, customer expectations, and other market effects, all of which may induce nonparticipants to buy high efficiency products.

The pretest was used to examine the question wording, and was used to identify 'early' signs of spillover. For the full-scale 2010 methodology being conducted in 2011, we have recommended that both like and nonlike customer spillover be measured as part of the free-ridership measurement to again identify early spillover, and approximately one year after measure installation to give respondents time to install efficient measures. We will coordinate the later measurement of spillover with the 2011 and 2012 non-energy impact (NEI) research being conducted with C&I customers.

Otherwise, the evaluation team made only minor revisions to the "like" spillover estimates for the pretest (as discussed above) and, at this point, no revisions to the nonparticipant spillover estimates. Due to the minor adjustments and the small role vendors played in this pre-test, nonparticipant spillover is not addressed in this analysis. Only savings estimates for "like" spillover are presented.

The evaluation team has included questions to address "unlike" spillover – energy efficient equipment installed by a participant due to program influence that is not identical to the equipment they received through the program. However, given the difficulties in estimating savings for these installations, we present only indicators of "unlike" spillover and not savings estimates. The joint NEI/spillover study to

be conducted later in 2011 and 2012 by expert interviewers will allow for better estimation of unlike spillover.

The spillover estimates are computed based on how much more of the same energy-efficient equipment the participant installed outside the program that were, in fact, influenced by the program. The following questions, in conjunction with the savings assigned to that same equipment by the program, are used to estimate possible spillover savings:

S1A Now I'd like you to think of the time since you participated in the <PROGRAM> in <DATE>. Has your company implemented any <MEASURE CATEGORY> projects on its own, without participating in a <PROGRAM ADMINISTRATOR> program, for this or other facilities in Massachusetts?

S1B Was this equipment of the same efficiency level or a higher level of efficiency as the equipment you installed through the program?

S2A About how many additional [IF EFF = 1: ENERGY EFFICIENT] <MEASURE CATEGORY> projects did your business implement on its own since participating in this program in 2010 compared to the amount you implemented through the program?

For respondents that answer “Yes” to S1A, spillover savings are calculated as the measure-specific savings identified by the program multiplied by the quantity identified in S2A. If the respondent answers “No” to S1A, there are no identifiable “like” spillover savings.

For those measures, a program-attributable spillover rate is then calculated based on the following questions:

S3A Did a recommendation by the contractor, engineer, or designer who you worked with under the <PROGRAM> influence your decision to implement some or all of this [IF EFF = 1: EFFICIENT] <MEASURE CATEGORY> project on your own?

S3B Did your experience with the energy efficient projects implemented through the <PROGRAM> influence your decision to implement some or all of this [IF EFF = 1: EFFICIENT] <MEASURE CATEGORY> projects on your own?

S3C Did your participation in any past program offered by <PROGRAM ADMINISTRATOR> influence your decision to implement some or all of this [IF EFF = 1: EFFICIENT] <MEASCAT > equipment on your own?

If the respondent reports that the contractor influenced their decision to install the like equipment on their own, we attribute the program with 50 percent of those savings based on the influence the program has on the trade allies. If the respondent reports that either their experience with the program-sponsored project or past programs influenced their decision to implement the like equipment, we attribute the program with 100 percent of the spillover savings.

To summarize:

If (S3A=yes AND (S3B = no AND S3C = no)), spillover rate = 50%.

If (S3B=yes OR S3C = yes), spillover rate = 100%.

That rate, applied to the estimated spillover savings, results in the program-attributable spillover savings for that participants. This method is very similar to the 2003 methodology. We recommend adding a 0-10 influence scale item as a consistency check to this series.

5.4.1 "Like" Spillover

Our pretest yielded minor amounts of "like" spillover. Only six of the 82 participants reported purchasing similar equipment to what was purchased through the program on their own. Of those six, only four purchased energy efficient equipment and of those four, only two reported that their experiences with the program influenced their decision. To identify spillover rates, we take the ratio of spillover savings to the savings of the sampled group. This analysis results in spillover rates described in Table 5-16.

Table 5-16. "Like" Spillover Rates by Program and Measure

Program	Measure	Sampled kWh savings	Spillover Savings	Spillover Rate
Energy Initiative	Lighting	3,743,748	236,764	6.3%
	Compressed Air	338,458	0	0%
Small Business Services	Lighting	917,306	1,277	0.1%

5.4.2 "Unlike" Spillover

In addition to "like" spillover, the 2010 methodology pretest also measured "unlike" spillover (i.e., end-uses outside of those installed through the program). Of the 82 participants interviewed, thirteen reported that they purchased additional equipment since they participated in the program. However, only two purchased equipment that would qualify for an existing Sponsor program. To establish spillover savings, program eligibility was used as a proxy for energy efficiency.

The two participants that purchased additional equipment purchased LED refrigerator lighting for four or five refrigerated cases and four kitchen exhaust hoods with variable frequency drives. Given the detail provided by these two participants, it would be possible to capture a rough estimate of savings that could be applied to the program.

The following questions were used to identify "unlike" spillover.

S5 Since participating in <PROGRAM>, had your company purchased, installed, or implemented any other type of energy efficient equipment on your own?

S6 What did you install?

S7 Would this project have qualified for an incentive through the <PROGRAM>?

S8A Why didn't you implement this project through a <PROGRAM ADMINISTRATOR> program?

S8B *[IF S8A = 4]* Why wouldn't the project qualify?

6. RECOMMENDED FINAL METHODOLOGY

This chapter summarizes our recommended free-ridership and spillover methodology to be used in the 2011 full-scale study with 2010 program participants. Section 3.1.2 of this report presented best practices for use in self-report free-ridership and spillover measurement, including survey design elements, data collection elements, and analytic elements. Although we have not repeated these elements in this chapter, the recommended final methodology follows these elements.

Based on the results of the survey pretest, we made minor adjustments to the wording in the pretest survey instruments. The algorithm remained unchanged. (The final data collection instruments and algorithms that incorporate these elements can be found in Appendix E and F, respectively.) The changes to the participant survey included the following:

1. Minor wording changes through the instrument to further clarify question wording. These changes are based on feedback from interviewers and monitoring of pretest interviews.
2. Added prompt for respondents about any interest-free financing they might have received as part of program participation. If the respondent received interest-free financing as part of the program, they are reminded of this assistance during the free-ridership battery.
3. Added prompts for "upgrading equipment" to R4b and R4c.
4. Moved C1 – influence of third party – to follow question regarding third party involvement (FR4).
5. Moved "Impact of Previous Program Participation" section to follow free-ridership battery.
6. Added three questions to assess the program's influence on unlike spillover.
7. Added consistency check scale items to both like and unlike spillover batteries.
8. Added questions regarding expected non-energy impacts.
9. Removed R4b and R4c. These questions were only used for process-related issues as part of the pretesting.
10. Added S1c (Was this equipment more energy efficient than standard efficiency or code equipment?) to assess partial spillover savings. If the efficiency level of the spillover equipment installed by the respondent was below what would qualify for the program but more efficient than standard or code equipment, the savings are reduced by 50 percent.

To the vendor survey, the following change was made:

1. Added pre and post participation sales volume questions to the vendor survey.

As discussed in Section 3, we recommend estimating free-ridership in a timeframe close to survey participation in order to reduce potential recall bias, and estimating spillover at a later date with these same participants to allow them time to install efficient measures outside the program. For the 2011 study, participation has already occurred in 2010 so more frequent measurement for free-ridership purposes is not possible this year, but should be used in future years.

For the 2011 study of 2010 participants, we have assumed two different data collection periods, one for the C&I electric study with a final report by mid-June and one for the C&I gas study with a final report by August 30, 2011. For the electric study, we have further assumed that we will provide separate free-ridership and early spillover estimates for the following large PAs having a sufficient number of participants—National Grid, NSTAR, WMECo, Unitil, and the Cape Light Compact. For the remaining electric PAs (and in cases where measure categories have less than ten participants), we will apply a statewide average. For the gas study, we will also apply statewide averages where the number of participants and/or measure categories are low.

Based on discussions with the PAs, it was also decided for this survey year to include the spillover questions in the free-ridership customer survey to measure 'early' spillover. The survey will also include

a few questions asking about anticipated non-energy impacts from their participation. A more comprehensive spillover effort will occur in the fall and in 2012 as part of the C&I non-energy impact interviews with a subset of the 2010 program participants interviewed in 2011 for the free-ridership and early spillover study. In each annual evaluation cycle, we will focus on measuring the NEI and spillover for a different segment of C&I participants. We will divide the C&I participant population into three categories based upon the type of program, specifically:

- 2011 – C&I Prescriptive Retrofit Programs
- 2012 – C&I New Construction Programs
- 2013 – C&I Custom Programs

Segmenting the customers by program type will allow us to capture the efficiencies in survey design and data analysis on an annual basis while achieving NEI estimates at the conclusion of the three year evaluation period.

APPENDIX A: TAXONOMY OF PA PROGRAMS BY SECTOR, TYPE OF ASSISTANCE, ELIGIBILITY, INCENTIVES, AND DELIVERY

In order to make recommendations on free-ridership/spillover methodologies that are best tailored to Massachusetts C&I program designs and needs, the evaluators attempted to interview C&I program managers/staff for all the PAs to better understand program delivery. In-depth telephone interviews were completed with C&I program managers/staff representing six of the PAs, with a total of seven PA staff persons and one implementation contractor representative participating in the interviews. Because these program managers/staff generally said they were not well enough informed to answer questions about past free-ridership methodologies and interpretation of results, we sought responses to these questions from utility EM&V representatives. The following matrix summarizes the responses from these program manager/staff interviews, as well as some additional information from utility EM&V representatives that was obtained via email.

C:\Documents and Settings\MAD

APPENDIX B: BEST PRACTICE REVIEW SOURCES OF INFORMATION

Cook, Gay, Summit Blue Canada Inc. *Attribution Methodology Wars: Self-Report Methods versus Statistical Number Crunching—Which Should Win?* Paper delivered at ACEEE. 2008.

Year(s) implemented	2008
Sponsoring agency/utility	ACEEE
Key Issues Addressed	Describes various approaches of assessing impact of free riders and spillover on program savings estimates, provides pros/cons of the methods, along with suggestions for determining which method is appropriate for certain types of evaluations.
Background	Three methods are commonly used to assess the NTG ratio: 1. Self-report methods: Survey of participants and nonparticipants what they would have done in absence of program support. *Enhanced self-report* methods involve calibrating information from other sources (interviews with vendors, trade allies, review of program documentation, analysis of market-based sales data, etc) with the survey results. The enhanced methods can also determine what additional efficiency improvements participants have made outside the program, how participating vendor sales practices would have been different without the program and how nonparticipating vendor and customer practices have changed since the program was implemented. • Pros: Simplest and lowest cost method is a telephone survey. Can triangulate different perspectives to measure correct construct with increased accuracy; directly addresses behaviors program attempts to affect; and flexible enough to take into account the program-participant interactions. • Cons: Can provide biased results; difficult to systematically convert opinions of participants into quantifiable free ridership values; limited participant recall; tends to underestimate spillover; virtually impossible to define a precision target and statistically valid sample size. 2. Econometric methods: Application of statistical tools and techniques to economic issues and data to develop models to compare participants' and nonparticipants' energy usage and demand patterns. Overall pros and cons: • Pros: Provides quantitative estimates of magnitude of net impacts; can provide more accuracy because tests for bias and precision can be included. • Cons: participants and nonparticipants included in a model; sample not randomly selected since participants are self-selected; no trade ally effects included. • Billing Analysis: Used to calculate annual demand and energy savings a. Pros: Can be used with complex retrofits and controls projects b. Cons: Large customers can have a significant influence; usable sample is reduced to customers with sufficient billing history; does not estimate spillover • Econometric models: Used to analyze co-relational relationships, usually with the hope of determining causation a. Pros: Can avoid concern over potential bias and cognitive dissonance issues with survey research; tests for bias and precision

can be included; can predict free ridership and spillover.

b. Cons: Inability to estimate spillover upstream in the distribution channel; robust study requires large budget for evaluation

- Discrete choice analysis models: Simulates the decision to purchase various types of commercial equipment, and then uses the model to determine the probability of purchasing high-efficiency equipment in the absence of the program.

3. Market share methods: Market shares approach uses aggregated sales volumes compared to baseline estimates of volume that would have been sold in absence of program. Saturation data analysis uses observations of the share of high efficiency equipment at two points in time.

   a. Pros: Assesses trends for the entire market; can estimate net energy impacts for programs where participation is not well defined.

   b. Cons: Collecting reliable saturation data requires a large budget and usually not repeated frequently; difficulty in collecting sales data (vendor concerns of releasing competitive data) and matching available data to unit of analysis (region, utility territory, etc).

Free Ridership Methodologies

Selection of method depends on:

- Objectives of program being evaluated
- Evaluation budget and resources
- Specific aspects of measure and program participants

For some programs, methodology selection is straightforward, as in the example below, with self-reported methods preferable.

Example Comparison of Methods for C&I Custom Programs:

Program Characteristic	Self-Report Methods	Statistical Models
Targets large customers.	In-person or telephone surveys can be used with large customers.	Large customers can overly bias results.
Non-participants difficult to identify.	Does not require non-participant data for free riders or inside spillover.	Requires both participants and non-participants in analysis.
May not detect savings at whole building/facility level.	Targets measure level information.	Energy use data generally only available at building/facility level.
External factors likely to be significant.	Survey accounts for relevant external factors.	Need to collect appropriate data to adjust for external factors.
Focused on process changes rather than equipment.	Survey accounts for changes to processes as well as equipment.	Discrete choice and other models focus on equipment choices.

Methods selection may be less straightforward, for example if the screening criteria point towards a market-based approach, but the market data is not available.

Conclusions

Market share approach is preferable when either there is not good data on participants and/or the goal is to assess market transformation. Surveys with participants and nonparticipants should be done as part of self-report methods or statistical methods.

Best approach includes following the guidelines for SR surveys in combination with interviews with other market actors, market share data, etc., to use the survey results in statistical modeling where appropriate and to select the approach that best meets the evaluation goal within the available resources.

A combination of methods may be used, with certain methods being used on an annual basis and another method being used at certain intervals.

Cooney, Kevin, Beth Baker, Timea Zentai, and Adam Knickelbein, Summit Blue Consulting, LLC, *Gas Furnace Market Transformation Model Development and Market Research.* Submitted to Energy Trust of Oregon. August 5, 2009. Presented by Fred Gordon at AESP Brownbag. 2010.

Year(s) implemented	2009
Sponsoring agency/utility	Energy Trust of Oregon
Sector	Residential: Single family, multifamily and manufactured homes
Key Issues Addressed	1. Develop a baseline estimate for the gas furnace market in the absence of the Energy Trust program 2. Adapt a current market transformation model for gas furnaces for use with other programs and increase the usability and functionality of the current model
Background	Several methods were used to develop the inputs to the model: • Secondary data review to facilitate more robust baseline and penetration values • Vendor interviews to help inform the baseline estimate. Trade ally contractors were selected that represented major portion of sales as well as ones who had not been as active. Distributors were selected based on territory. Additional distributor survey collected data on unit sales for the county. • Researching savings associated with federal code changes to assess relationship of utility program to savings achieved through the new standards in their territory. Due to the uncertainty of some of the inputs of the model, two baselines were developed: • *Low case baseline*: Compilation of interviews within county plus national market data • *High case baseline*: National market data alone Two factors change for each scenario: the baseline and number of gas conversions in the service area. Because the high case uses a lower baseline and higher number of conversions, the results show higher attributable energy savings. The model considers only the retrofit market and the units incented under the program, but the baselines penetration values do not differentiate between the retrofit market and new construction market. The market transformation model was originally created in Excel and was updated with user interface enhancements, functionality and updated assumptions and baseline values. There is neither precise data nor perfect sample and it is unlikely there will ever be. Indicators of market change include: • Multiple sources agree that market share is high • Nonparticipant vendors show high market share • Surrounding territories show high market share • Data over several years shows high market share • A small fraction on sales use program incentive.

	Sources of information include distributor interviews, contractor interviews, customer free rider questions, studies of nearby areas, parties to federal standards agreement. Keys to agreement of market change include: • Sustained effects • Complimentary sources of data pointing in same direction • Both market tracking and causal evidence • Open process and ongoing discussions, acknowledge uncertainty • Input of trade allies and evaluation experts and other key stakeholders • Pick a middle-lower number out of a range of possible savings numbers Only claim to accelerate for a few years.
Conclusions	NEAA has been doing for a while—occasional market studies, and then annual vendor surveys.

Fagan, Jennifer, Mike Messenger, and Mike Rufo, Itron, Inc. and Peter Lai, CPUC Energy Division.
A Meta-Analysis of Net-to-Gross Estimates in California. AESP. 2009.

Year(s)	2009
Sponsoring agency/utility	AESP
Key Issues Addressed	This paper provides an overview of estimates of the proportion of free riders in CA, reviews the pros and cons of different net estimation methods for free ridership only and methods used to estimate net market effects that include participant and nonparticipant effects.
Background	The authors claim that understanding net is important for: 1. understanding program and portfolio cost-effectiveness 2. improving portfolio design and resource allocation 3. refining program design and tactics 4. understanding market transformation 5. aligning program administrators' financial interests with societal interests 6. understanding how energy efficiency programs affect baseline load forecasts and short-term power procurement decisions.
Methodology	Seven different methods used to estimate NTGR, and their challenges: 1. Customer self-reports. Use for more traditional downstream programs. Challenge for upstream programs is that product buy-downs or instant discounts make the program invisible by design to many customers. 2. Supplier self-reports. Manufacturer/retailers' predictions of product sales with and without the program rebates often been used (especially for upstream rebate programs). Supplier responses may be biased because they realize that giving the right answers can effect continuation of program incentives. 3. Sales based method—per-capita sales comparisons with a comparable state that does not have a program (representing baseline sales). Comparison states must be very similar to program area or normalized statistically for differences in customer and market characteristics. Numerous limitations to this method--available adoption data is not always reliable, nor are data on the necessary normalizing variables always available. 4. Sales based method—paired comparison approach (used in Wisconsin). Comparing energy efficiency product sales data for a leading big box retailer in a state with rebates vs. a similar nearby state without rebates. Challenge is to find a representative retailer and controlling for differing demographics. 5. Econometric—discrete choice analysis. Estimates efficient product purchases as a function of factors that influence energy efficiency demand such as product awareness, prices, and other factors. Can be difficult in rapidly transforming markets where prices and product content are dynamic. Hasn't been used extensively to estimate net impacts due to complexity and expense. It relies on a large body of nonparticipant survey data (usually 3,000+). 6. Econometric—estimating a demand model. This model predicts the relationship between changes in energy efficient product price, different levels of customer awareness, and energy efficient product sales in

	different regions of the country (e.g., CA CFL study). 7. Econometric--net billing analysis. Can be used for measures that account for a minimum of 5-10% of end use consumption. Not useful where measures and savings in question are very site specific (e.g., industrial customers).
	The NTG estimation method to use depends on the specific circumstances and goals of the evaluation. Answering the following questions can help guide the choice: • What are the policy goals of the program? For short term resource acquisition, quantifying nonparticipant spillover and broader market effects may not be of interest. Where market effects are a strong objective, methods 3, 4, and 6 above are usually used. • How mature is the program? If program is in infancy and measures are less well-known, a sales based approach is not useful. Mature programs that have been successful and run for at least three years are likely to have some market effects and warrant use of sales based approach. Exception is for products promoted in neighboring jurisdictions for years that may be new to your program (e.g., CFLs). • What is the program design? Upstream or downstream. If upstream, method 2 may not work due to sample bias. If program promotes customized measures only, the econometric and sales based approaches are not feasible, and for industrial billing analysis can't be used. • How much budget do I have? Methods 7, 3, and 4 are less costly than Method 5. • What data is available? Sales based approaches (method 4) rely heavily on publicly available data sources for information. These data are often incomplete. If a good, complete and reliable data source is available, then this approach may be the best choice for assessing the full influence (FR+SO). • Is a suitable comparison group available? Method 4 relies entirely on finding a representative retailer operating in jurisdictions with and without rebates. Methods 3 and 6 also require a good deal of diversity in the market conditions and the availability of non program areas. • What level of precision is desired? Should be based on the needs of policy makers and the program. If need high level but budget is limited, may rule out use of multiple methods or more costly methods. • Are there performance-based metrics to be met? If there is sufficient budget and available data to support use of more than one method, it is best to use multiple approaches and triangulate the results.
Analysis	Their review found that despite widespread changes in markets and multitude of NTG methodologies, in general, portfolio-level NTG ratios have been relatively constant since 1980.

Conclusions	Authors note that challenges associated with measuring net effects of energy efficiency programs are no more daunting than those facing other professions (e.g., education, public health, pharmaceutical, other policy and medical interventions). A wide range of NTG estimates can be derived from the same baseline data depending on NTG definitions, analysis methods, and time frame (immediate past or forecast of near-term future). Jurisdictions should cooperate in the collection of sales and market share data for efficiency products to expand available data and reduce evaluation costs. Need for oversight agencies to plan for potential market effects of programs operated over long periods of time.

Goldberg, Miriam L., J. Ryan Barry, Tammy Kuiken, Ben Jones, Paulo Tanimoto, Nicole Buccitelli,
Colin Rickert, and Darcy DeAngelo-Woolsey; KEMA, Inc., *Business Programs: Acceleration Treatment and Life Cycle Net Savings.* Submitted to the Public Service Commission of Wisconsin. March 10, 2010.

Year(s) implemented	2010
Sponsoring agency/utility	Public Service Commission (PSC) of Wisconsin
Goal	1. To review methods utilized by other jurisdictions and investigate the effect of acceleration on Focus attribution results by employing other methods 2. To investigate the effects of using life cycle net savings (LCNS) assumptions on the Focus attribution results
Key Issues Addressed	The intent of the effects of acceleration treatment analysis was to clarify how much of the difference between Focus and other programs' NTG ratios may be due to differences n the treatment of acceleration when determining program attribution..
Effects of Acceleration Treatment	The study reviewed the attribution methodologies of well-established, large-scale, nonresidential programs in California, Massachusetts, New York, Oregon and Vermont.

Comparison of Focus Method to Other Jurisdictions

State	Primary Treatment of Acceleration	Primary Data Collection Technique
Wisconsin Focus Y1NS	Acceleration less than 48 months receives partial credit towards attribution. Acceleration 48 months or more receives full attribution.	Self report participant surveys
California[1]	Acceleration less than 6 months receives no acceleration credit. Acceleration more than 6 months, but less than 48 months receives partial credit towards attribution. Acceleration 48 months or more receives full attribution.	Self report participant surveys
Massachusetts[2]	Acceleration more than 12 months receives full attribution. No partial acceleration credit given for less than 12 months.	Self report participant surveys
New York[3]	Acceleration less than 60 months receives partial credit toward attribution. Acceleration 60 months or more receives full attribution.	Self report participant surveys
Oregon[4,5]	The evaluation uses the program's effect on timing (yes/no) in developing the scores used to determine attribution. The length of the acceleration period is not considered.	Self report participant surveys
Vermont[6]	The most recent Efficiency Vermont Program C&I impact evaluation did	N/A

	not attempt to assess attribution.	

The various methods of acceleration were applied to the Focus evaluation. Results of this analysis indicated that the current Focus evaluation treatment of acceleration provides attribution results comparable to those in other jurisdictions. Final attribution scores are not highly dependent on the acceleration calculation methodology.

The effects of efficiency and acceleration on attribution are relatively equal: removing partial credit for either causes attribution to decline by roughly 10% versus the current Focus evaluation method.

The study examined why varying the acceleration treatment had such a limited effect upon attribution. The attribution scores were grouped into categories:

- None: an attribution score of zero
- Partial: an attribution score between zero and one
- Full: an attribution score of one
- Market-based: an attribution score determined by a market study

Varying the acceleration method only affected the net savings for measures that received partial attribution scores. The measures with no attribution or market-based attribution scores were not affected by changes to the acceleration treatment. The majority of measures with full attribution would not be affected by changing the acceleration method, except for ones with acceleration between 48 and 60 months. For these, the attribution scores would be reduced using the NY acceleration method, but would remain the same under any of the other acceleration methods.

Life Cycle Net Savings (LCNS) Approach

The LCNS method provides for a different treatment of accelerated projects and produces lifetime net savings instead of the first year net savings that the current Focus method employs. Savings in the LCNS method are based partly on length of time a measure remains operational, so measure life is a key input to this method. The LCNS method does not incorporate a discount rate such as what would be included in a full benefit/cost analysis.

Similar to the 1st-year method, LCNS calculates attribution as a ratio of net savings to a ratio of verified gross savings, but has two differences:

1. LCNS looks at the total lifetime savings of the equipment
2. LCNS increases the annual verified gross savings in the acceleration period for custom measures where the existing equipment had lower than standard efficiency. In the post-acceleration period and for non-accelerated measures, the annual verified gross savings are the same as those used in the 1st-year method. The ratio of the two saivings is referred to as the A/P ratio (Annual savings in acceleration period divided by savings in the post-acceleration period).

For some measures, the annual gross savings had to be estimated since the input data needed to calculate annual gross savings was not available.

The study used two different A/P ratios to investigate the uncertainty in the assumption of the A/P ratio and to confirm the robustness of the results. The table below shows the differences between the 1st-year savings method and the two methods of LCNS tested in this study.

Assumption	Y1NS	LCNS Method A	LCNS Method B
Type of savings	First year savings	Lifetime savings	
Annual acceleration period verified gross savings	The difference between the energy use of the rebated equipment and the energy use of its standard efficiency replacement.	The difference between the energy use of the rebated equipment and the energy use of the equipment replaced.	
Annual post-acceleration period verified gross savings		The difference between the energy use of the rebated equipment and the energy use of its standard efficiency replacement.	
Acceleration period net savings	n/a	Acceleration period verified gross savings multiplied by the acceleration period.	
Post-acceleration period net savings	n/a	Post-acceleration period verified gross savings times the simple program attribution (SPA).	
A/P ratio assumed for custom CATI	n/a (implied 1)	2	Based on sector level A/P ratios observed in the engineering survey
Net savings calculation	Verified gross savings times [SPA + (Acc/48)(1-SPA)]	Acceleration period net savings plus post-acceleration period net savings	

Both LCNS methods resulted in lower attribution factors than those calculated using the 1st-year method. The study found the difference was less about the acceleration treatment than the difference between weighting measure attribution by 1st-year versus lifetime savings. The 1st-year method results in a higher savings for shorter-lived measures than on measures with longer lifetimes.

The lower attribution from the LCNS method was apparent across all sectors, but the largest difference was in the Agriculture and Commercial sectors. These sectors had a large amount of savings from CFLs, which receive high market-based attribution scores. A shorter measure life, such as with CFLs, results in less lifetime savings than measures with similar savings with longer lifetimes.

The two methods of LCNS tested resulted in similar attribution factors, when the attributions were rounded to the nearest percent. Custom CATI measures accounted for only a small portion of savings, so the A/P ratio had a limited ability to affect the results.

The paper concludes with a recommendation to the PSCW to consider further development and refinement of the LCNS method. The two main differences between the approaches:

1. The first-year approach treats the reported acceleration period more as an indicator of the likelihood the measure would have been installed without the program rather than as a literal indicator of the time until the measure would have been installed.
2. The first-year approach determines aggregate attribution for a program, sector, or portfolio weighting measures only by first-year savings. The life cycle approach weights measures according to lifetime savings. The first-year approach gives more weight to shorter-lived measures

[1] Nonresidential Net-to-Gross Working Group. *Methodological Framework for Using the Self-report Approach to Estimating Net-to-Gross Ratios for Nonresidential Customers.* February 9, 2009.

[2] Sponsor utilities included National Grid, NSTAR Electric, Northeast Utilities, Unitil, and Cape Light Compact.

[3] NYSERDA. *Annual Report for 2008 – Program Evaluation and Status Report – Issued March 2009*, Section 2.3 Largest Savers Impact Evaluation. December 31, 2008. http://www.nyserda.org/publications/default.asp.

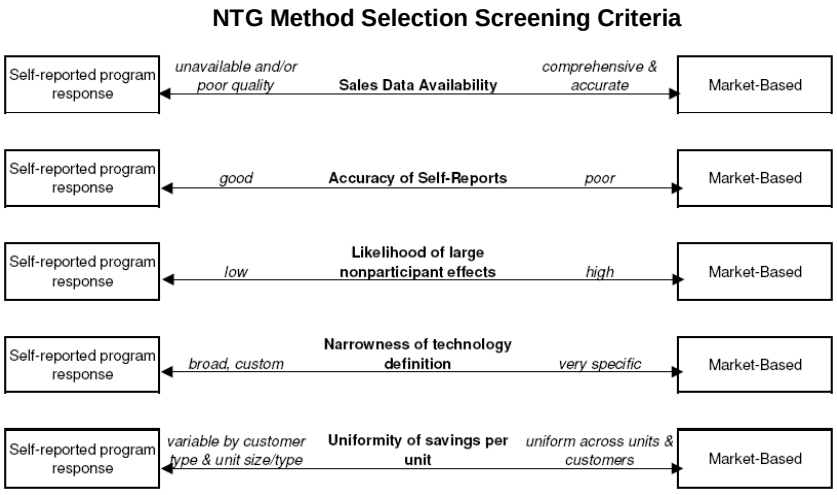
[4] Energy Trust of Oregon, Inc. *Evaluation Committee Report.* May 11, 2007. http://www.energytrust.org/meetings/board/2007/070808/04a_EvalMay.pdf.

[5] ADM Associates, Inc. *Impact Evaluation of New Building Efficiency Program for 2004 and 2005, Final Report.* February 2008.

[6] KEMA, Inc. and RLW Analytics. Final Report: Phase 2 Evaluation of the Efficiency Vermont Business Program. February 2006. http://publicservice.vermont.gov/pub/other/evaluationoftheefficiencyvtbusprogfinalreportphase2.pdf.

Goldberg, Miriam, KEMA Inc., Oscar Bloch, Wisconsin Department of Administration, Ralph Prahl, Prahl & Associates, David Sumi and Bryan Ward, PA Consulting Group, and Rick Winch and Tom Talerico, Glacier Consulting Group. *Net-to-Gross Method Selection Framework for Evaluating Focus on Energy Programs.* Prepared for Public Service Commission of Wisconsin. March 16, 2006

Year(s) implemented	2006
Sponsoring agency/utility	Public Service Commission (PSC) of Wisconsin
Sector	Residential, Business Programs (Agriculture, Commercial, Industrial, Schools & Government), Renewable Energy Programs
Key Issues Addressed	Provide a method selection framework to make the rationale for the choice of methods more transparent to users of the evaluations and to provide a greater confidence in the results.
Background	In 2006, the Focus on Energy (FOE) Evaluation team and the PSC developed a decision tree to guide evaluators in deciding whether a self-reported program response method or market sales-based method would be more appropriate for the evaluations. Three steps are involved in choosing a method: 1. Definition of measure groups to be analyzed separately 2. Determining the best net-to-gross (NTG) method for each group 3. Determining the detailed data collection and analysis methods for the NTG method for each group. The method choice is based on the following considerations: 1. Sales data availability: Current and baseline market sales data 2. Accuracy of self-reports: Ability of end-users and/or vendors to report accurately what would have occurred in the absence of the program 3. Likelihood of large nonparticipant market effects: likelihood of substantial nonparticipant market effects, indicating need for methods to capture such effects 4. Narrowness of technology definition: whether the technology to be addressed by a single analysis effort is a single technology or multiple categories of technologies. 5. Uniformity of unit savings: Whether the savings per unit is sufficiently consistent across types of units and customers that the program effect can be adequately quantified in terms of the total number of units sold, rather than requiring information on unit characteristics by customer type.

	
Methodology	Different methods can be chosen for different measures within a program area or program. Method selection is an iterative process based on the way that the groups to be analyzed are defined. The first step is defining groups to be analyzed separately, i.e. single measure in a broad market, entire program spanning a broad set of measures in a broad market, etc. Once a method is or methods are assigned for those groups, the groups may be combined to analyze together or split into smaller groups if it is anticipated that different subgroups would lead to different answers.
Eligible respondents	Market sales-based methods: Relies on aggregate data on total sales of specific technology in WI Self-reported Program Response: End-users and/or vendors
Types of measures	Method can be applied to specific technologies or defined groups of technologies

Itron, Inc. and KEMA. *2004/2005 Statewide Express Efficiency and Upstream HVAC Program Impact Evaluation*. December 31, 2008.

Year(s) implemented	2004-2005
Sponsoring agency/utility	CPUC
Sector	Commercial retrofit
Goal	Encourage the installation of select high efficiency equipment.
Timing of measurement	Unknown
Eligible respondents	Participants and nonparticipants
Types of measures	Lighting, HVAC, refrigeration, and motors.
Free ridership questions for customers (downstream self-report)	Two approaches were used to estimate free-ridership, and the two resulting estimates were averaged. **Three-criteria approach** consisted of three questions: 1. LI42: If the rebate or cash incentive did not exist, which of the following best describes what you would have purchased… • You would NOT have purchased new equipment • You would have purchased fewer new equipment or less new equipment • You would have purchased the same quantity of equipment as you did through the program 2. LI43: If the rebate or cash incentive did not exist, which of the following best describes what you would have purchased… • Standard efficiency equipment or the least expensive alternative available • Less efficient than the equipment we just discussed • The same high efficiency equipment as you purchased through the program 3. LI44: If the rebate or cash incentive did not exist, would you have installed the rebated lighting equipment… • More than 1 year later • Within 1 year • At the same time **Program-influence approach** consisted of one question: On a scale of 1-10, with 1 being *NOT AT ALL* Influential and 10 being *EXTREMELY* Influential, how influential was the Express Efficiency program rebate or cash incentive on your decision to install the rebated equipment?
Free ridership algorithm (downstream self report)	**Three-criteria**--If the respondents states that he or she would have purchased the same quantity and type of equipment, at the same time, and at the same level of efficiency, the are scored as a free rider. Likewise, if the respondent states they would not have purchased the equipment or would have purchased standard equipment, they are scored as a non-free ridership. Partial free ridership is scored based on the frequency (0-3) of partial free ridership responses (e.g., would have purchased less new equipment). Self-report and discrete choice methods are used to evaluate CFLs, T8s, and AC systems. For CFLs and AC system, the self-report method is used as the

	evaluators speculate that the discrete choice methodology is not accounting for upstream program effects. **Program-influence** --Free ridership was calculated directly from this response, with a 1 indicating a customer was a free rider (FR = 1.00) and 10 indicating a customer was a net participant (FR = 0.00). All other values of free ridership were interpolated between these two points using the following equation: Free Ridership = 1−(influence rating −1)9 The average of these two results was taken as the final free-ridership estimate.
Free ridership approach (upstream program)	An upstream approach was utilized the calculation of NTFR ratios for motors and central air conditioners (CAC). Using in-depth interviews and CATI surveys, participating motor and CAC distributors were asked: • "What proportion of the rebated <SPECIFIC CAC/MOTOR MEASURE> you sold in 2004 and 2005 do you think you would have sold in California if you hadn't participated in the program?" This was followed by a confirmation question which read, "Okay, just to confirm – you are saying that <PROPORTION STATED> < SPECIFIC CAC/MOTOR MEASURE> would have been sold anyway in California if the program rebates were not available in 2004 and 2005. Is this correct?" Evaluators asked motors distributors to provide free ridership estimates for four different motor size categories and the CAC distributors to provide estimates for five different CAC size/efficiency categories.
Discrete choice modeling	A discrete choice modeling methodology was used to estimate a net of free ridership model for the non-residential audit-only, rebate program-only and combined-program net of free-ridership ratio for lighting and HVAC equipment measures.
Spillover questions for customers	Not assessed.
Spillover questions for vendors	Not assessed.
Spillover algorithm	NA

Keating, Kenneth M., PhD. *Free-Ridership Borscht: Don't Salt the Soup.* IEPEC. 2009.

Year(s) implemented	2009
Sponsoring agency/utility	IEPEC
Key Issues Addressed	This paper documents the inadvertent bias that multiplicative algorithms can introduce into net-to-gross estimates.
Background	In most energy efficiency evaluations, free ridership receives the most attention as stakeholders want to avoid spending ratepayer or public funds on measures or behaviors that would have occurred without those funds. The self-report approach (asking the participants a set of related questions to try determine their motivation) is a popular methodology due to its direct approach and transparent nature. However, when converting the set of responses from participants into a probability of free ridership, evaluations can advertently bias results in one direction by multiplying individual measurements scores together into one summary score.
Free Ridership Methodology	Rather than using multiplication, employ averages of individual but similarly scaled measurements or averages of macro indices (such as a four question series of questions). Multiplication is acceptable if only conditional probabilities are factored and each probability is independent of the others (e.g., efficiency levels cannot be included in timing or quantity probabilities) or if applied to actual savings values (e.g., kWh saved). This type of self-report protocol is typically used in residential or small commercial evaluations. Large commercial or industrial evaluation requires more detailed inquires as efficiency levels can be assessed easily with a scaled response.
Conclusions	Many of the criticisms of the self-report approach can be mitigated through careful research design, sampling, survey timing, and question wording. Avoid the use multiplication except in carefully worded questions that assess conditional probabilities or when applied to a real world value such as gross savings.

Keneipp, Marshall, Floyd Keneipp, and Jeff Erickson, Summit Blue Consulting, LLC and Bill Norton, Opinion Dynamics Corp. *APS Measurement, Evaluation, & Research (MER) Report, Consumer Products Program (CPP).* APS. September 30, 2008.

Year(s) implemented	August 2005 through December 2007
Sponsoring agency/utility	APS
Sector	Residential
Goal	Program promotes the purchase of high-efficiency ENERGY STAR-rated CFLs through discounted pricing at participating retail outlets
Timing of Measurement	Surveys conducted in December 2006 and September 2007
Eligible respondents	APS customers who purchased CFLs
Type of measures	CFL bulbs
Free ridership methodology	Participants were asked the same questions for 2 types of bulbs in a phone survey, with sample data including type of CFL purchased, the amount paid , date of purchase, and store where purchased: FR1 Would you have paid up to $[price paid + buydown amount + 50% of buydown] for [desc of product]? *(if yes, skip to 4th question)* FR2 Would you have paid up to $[price paid + buydown amount]? *(if yes, skip to 4th question)* FR3 Would you have paid up to $[price paid - 50% less than buydown] for [desc of product]? FR4 if the [type] CFLs you purchased at [store] on [date] had cost $[price paid +buydown amount] would you have purchased: • More CFLs [+0%] • Definitely the same number [+0%] • Probably the same number [-10%] • Probably fewer [-25%] • Definitely fewer [-50%] • Don't know/don't recall [-0%] 10. Prior to purchasing these bulbs were you… 1.Not at all familiar with CFL bulbs (also called CFLs) [skip to 11) 2.slightly familiar with CFLs 3.Somewhat familiar with CFLs 4.Very familiar with CFLs 5.(don't know/refused) 10b. Prior to purchasing these bulbs, would you say that you used … 1. No CFLs (0%) 2. Some CFLs 3. About half CFLs 4. Mostly CFLs 5. all CFLs in the screw-in sockets in your home (100%)

	6. (don't know/refused) 14. Prior to purchasing the bulbs at [store] on [date], had you purchased any CFL bulbs? 1. Yes, I have purchased CFLs before 2. No, this was my first CFL purchase 3. (don't' know/don't recall)
Free ridership questions for vendors	None
Free ridership algorithm	A participant is initially defined as a 100% free rider if they would have bought the CFLs at an unsubsidized price (by answering yes to either question FR1 or FR2). 100% free riders where then asked question FR4 and their free ridership percentage was adjusted by the percent shown next to the response of question FR4 above. The NTG survey results were weighted according to the number of CFLs purchased by the respondents to give a savings weighted NTG estimate. Summing gross and net Watts across the surveyed population then dividing the net Watts by the gross Watts gives the final savings-weighted NTG ratio. One minus the NTG ratio is the free ridership percentage. The initial free ridership rate was discounted based on answers to questions 10, 10b and 14. the discounted free ridership rate was calculated assuming the following were NOT free riders: • Someone "not at all familiar" with CFLs • Someone that had used no CFLs before • Someone that had purchase no CFLs before The free ridership total was also examined if someone "slightly familiar" with CFLs was considered not a free rider, but that result is less defensible.
Spillover questions for customers	The participant survey included the following questions: SO1. Have you purchased any additional CFLs since the purchase that we've been discussing? SO2. [if yes to SO1] Did you receive a discount or did you buy these additional CFLs at a reduced price? SO3. Would you have purchased these additional CFLs if you did not have the prior experience of using the CFLs that we've been discussing? SO4. How many CFL bulbs have you purchased since [date]? SO5. To the best of your knowledge, did the information you received from APS in any way influence your decision to purchase these CFLs?
Spillover questions for vendors	None
Spillover algorithm	Questions SO3 [No] and SO5 [Yes] define who purchased CFLs that ought to be counted as spillover for the number in SO4. To extrapolate to the population, the average CFL Watts for respondents were used as a proxy for spillover bulbs. If the number of bulbs was not known, the average of those who did know was used. Summing spillover Watts across the respondents and dividing by gross reported Watts for all respondents yielded spillover as a percent of total reported savings.

Klos, Mary and Joan Huston, Summit Blue Consulting, LLC. *Impact Evaluation of 2007 CFL Buy-Down Pilot.* Prepared for Progress Energy—Carolinas. May 20, 2008

Year(s) implemented	2007-2008
Sponsoring agency/utility	Progress Energy—Carolinas (PEC)
Sector	Residential
Goal	Pilot program to increase consumer awareness of benefits of ENERGY STAR CFLs by providing educational material and a discounted bulb price to consumers.
Timing of Measurement	2 surveys – one shortly after purchase, one four months later
Eligible respondents	Purchasers of CFL multi-packs
Type of measures	CFL bulbs in multi-packs
Free ridership questions for customers	Three questions from the surveys were used to assess free-ridership: • Thinking about the price of the CFL bulbs last fall at Home Depot, on a scale from 1 to 5, where 5 means "very important" and 1 means "not important at all," how important was the sales price in your decision to buy CFL bulbs at that time? (follow-up survey) • Do you already have any CFL bulbs like these in your home? (initial survey) • How many CFL bulbs are you already using in your home an in which rooms are you using them? (initial survey)
Free ridership data from vendors	Collected CFL sales levels at Home Depot 9 weeks before and during the pilot.
Free ridership algorithm	Looked at free-ridership in three ways: 1. Responses indicating the sales price was of little or no importance in their decision to purchase the bulbs were viewed as free-riders (as a percentage of respondents). 2. Respondents who previously had installed CFLs were considered free-riders (percentage of respondents). [Not quantifiable, but low based on large increase in number of CFL bulbs installed per home.] 3. Level of product sales (bulbs per week) recorded pre-program were considered the level of free-ridership during the program. The midpoint between the first and the third measures was used as the best estimate for free-ridership.
Spillover questions for customers	Three questions addressed spillover: • Have you purchased additional CFL bulbs for your home since the special sales price …? • How likely are you to purchase additional CFL bulbs for your home in the future? (1=Very unlikely, 4=Very likely) • Based on your experience with CFL bulbs, how likely would you be to recommend them to family or friends? (1=Very unlikely, 4=Very likely)
Spillover data from vendors	Collected CFL sales data for before, during and after buy-down pilot
Spillover algorithm	Four indicators were examined: 1. For people that said they purchased additional bulbs, deduced

	how many additional bulbs were purchased by combining rate of installation, total number of CFLs installed and number of people who said they purchased additional bulbs to calculate average purchase of bulbs per customer who bought additional bulbs. Estimated spillover rate from this number. 2. Percentage of customers very likely to purchase additional bulbs in the future 3. Percentage of customers very likely to recommend CFLs to family or friends 4. Comparison of bulb sales before, during and after buy-down pilot. The increase in sales after the event is considered the spillover effect. Indicators 2 & 3 were not easily quantifiable to spillover estimate, but high numbers lend support to substantial spillover. Indicators 1 & 4 were quantifiable, so a mean of the two was used as the spillover estimate for evaluation purposes.

Megdal, Lori, Megdal & Associates, LLC, Yogesh Patil, Energy & Resource Solutions, Inc., Cherie Gregoire and Jennifer Meissner, New York State Energy Research and Development Authority, and Kathryn Parlin, West Hill Energy & Computing, Inc. *Feasting at the Ultimate Enhanced Free-Ridership Salad Bar*. IEPEC. 2009

Year(s) implemented	2005-2007
Sponsoring agency/utility	NYSERDA (2009 IEPEC conference)
Key Issues Addressed	While unsophisticated batteries of questions with arbitrary scoring will generate NTG ratios at little cost, complex projects require a "salad bar" approach using independent review, mixed modes, and multiple viewpoints. This methodology provides results that "demonstrate construct validity, consistency, and low variation."
Background	Because net-to-gross ratios were a major point of uncertainty for NYSERDA's large C&I evaluations, the researchers designed an evaluation of 25 of the largest savers in NYSERDA's programs. These programs included: • CIPP (Commercial/Industrial Performance Program) • DG-CHP (Distributed Generation – Combined Heat and Power) • NCP (New Construction Program). • PLMP (Peak Load Management Program) • TA (Technical Assistance) These large savers represented 18 percent of the incremental savings for the entire portfolio.
Free ridership and spillover methodology	The researchers employed a "salad bar" approach where all respondents received a core set of questions, and select instruments were applied to specific projects as appropriate. This selection was often determined by the decision-making process at each project. An initial telephone interview was conducted to determine the decision-makers for each measure at each site and obtain contact information for those decision-makers. Then, follow-up instruments were administered to the decision-makers either via telephone or during in-person interviews. These interviews yielded a wealth of quantitative and qualitative data. This data was then independently reviewed by three senior evaluation analysts. Each analyst determined a free ridership and spillover score for each measure and then teleconferenced with the other analysts to determine a consensus score. These scores included a range of values representing the uncertainty and potential measurement error of using both qualitative and quantitative data.
Response to criticisms of SRA	The researchers speculate that any social desirability bias found in energy efficiency self-report has limited effect on net-to-gross as research in more socially sensitive errors (drug abuse, sexuality) has shown small underreporting biases with little overall effect. Using data from multiple decision-makers limits the effect of any self-report bias or measurement error from one respondent. An in-depth methodology with multiple sources also allows evaluators to weight the value of responses from different decision makers. Using a large variety of free ridership questions allows for a detailed comparison of responses across questions. This comparison can enhance and test the consistency of the responses across questions.

Conclusions	When evaluating the decision making process at large projects (greater than 1.5 GWh in expected savings), using customized, site-specific methods leads to highly defensible and consistent results. However, this paper did not discuss the costs associated with this methodology – both financial and in terms of respondent burden. In addition, this methodology only presented an estimated range of free ridership; not a point value that is required in many regulatory environments.
Evaluation(s) using method(s)	NYSERDA Large C&I

National Action Plan for Energy Efficiency (2007). *Model Energy Efficiency Program Impact Evaluation Guide.* Prepared by Diane Munns and Jim Rogers. <www.epa.gov/eeactionplan>

Year(s) implemented	2007
Sponsoring agency/utility	EPA
Key Issues Addressed	Chapter 5 defines net savings, the four key factors that differentiate net and gross savings, provides descriptions of several approaches for determining net savings, and discusses criteria for selecting an appropriate net savings approach.
Background	There are three primary factors that differentiate gross and net savings: free ridership, spillover, and rebound. Free ridership is the most commonly evaluated net-to-gross factor (NTGR), then spillover, then rebound. Free riders are participants who would have taken the same action in the absence of the program. The program can also influence the timing, the level of efficiency, and the number of units installed. These different levels of free ridership are referred to as partial or deferred free riders. A non free rider would not have installed the baseline measure without the program. Free ridership can vary from one measure to the next and over time. Free ridership is a source of energy and demand savings uncertainty. Spillover occurs when there are demand or consumption reductions because of the program, but the program doesn't directly influence the behavior. This may occur because of additional actions participants take outside the program as a result of their participation, changes in the mix of equipment that manufacturers, dealers and contractors offer all customers as a result of program availability, changes in specification practices of architects and engineers, and direct or indirect changes in energy use of nonparticipants as a result of the program (e.g., advertising, stocking practices, changes in buying habits). Estimating spillover and free ridership is complicated by market noise, making it difficult to estimate a program's influence. Rebound occurs when participants increase their use of the equipment as a result of it's improved efficiency. Could argue that there is a non-energy benefit associated with increased comfort, health, and safety.
Free Ridership methodology	Chapter 5 discusses four approaches for determining the NTGR. All four approaches can be used with any type of program (assuming a large number of participants for econometric). 1. Self-reporting surveys, which use survey-based stated intentions from participants and nonparticipants. The best use of self-reporting surveys involves asking a series of questions. Responses to questions are combined (additively or multiplicatively) into an individual free rider estimate. While this is the lowest cost approach, it has disadvantages such as potential bias and overall accuracy. Using techniques like adding consistency check questions can improve survey quality. 2. Enhanced self-reporting surveys, which combine interviews and other data sources. For example, interviews with multiple decision makers (e.g., managers, engineers, facilities staff, contractors, design engineers, manufacturers, distributors, retailers). Another data source is a project analysis which looks at how the project addresses barriers and/or documentation the participant may have of the decision to proceed (e.g., memos, feasibility studies). Other data sources include market sales data, review of similar programs, market potential or effects studies. 3. Econometric models to compare participant and nonparticipant

	energy and demand. Can only be used with programs having a large number of participants and a comparable nonparticipant group. Also, the program must be large enough to justify the cost of this type of analysis. 4. Stipulated net-to-gross ratios, which are multiplied by gross savings to obtain net savings. Typically stipulated by regulatory body when the expense and uncertainty of the results are significant barriers. Due to the cost of estimating NTGR, it is acceptable to perform NTGR analyses less frequently (e.g., every few years) than gross savings impact evaluation as long as no major changes in market or behaviors.

Nonresidential Net-To-Gross Ratio Working Group. *Methodological Framework for Using the Self-Report Approach to Estimating Net-to-Gross Ratios for Nonresidential Customers.* May 8, 2009.

Year(s) implemented	For the evaluation of 2006-2008 programs
Sponsoring agency/utility	CA Public Utilities Commission (CPUC)
Sector(s)	Large nonresidential
Goal	Provide a standard framework using the Self-Report Approach (SRA) to estimate project and program-level Net-to-Gross Ratios (NTGR). The framework includes decision rules for systematically and consistently integrating findings from both qualitative and quantitative information.
Key Issues Addressed	• The method uses 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.
Background	A working group was formed as part of the evaluation of the 2006-2008 programs, tasked with developing a standard methodological framework for estimating net-to-gross ratios. This approach was 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). The method is a general framework that can be customized for individual programs. The approach has three levels of free-ridership analysis based on project type. The evaluators determine which projects are assigned to each category. The categories are described as: 1. **Standard – Very Large Project** NTGR: Most detailed level of analysis applied to the largest and most complex projects with the greatest expected levels of gross savings. 2. **Standard** NTGR: Somewhat less detailed level of analysis applied to projects with moderately high levels of gross savings. 3. **Basic** NTGR: Applied to all remaining projects. Each level of analysis relies on up to five sources of information: 1. Program Files: Includes documentation such as completed application form(s), correspondence between customer and utility representatives, notes on project details, copies of rebate checks, etc. 2. Decision-Maker surveys: A survey is conducted with project decision makers to obtain highly structured responses concerning the probability that the customer would have implemented the same measure in the absence of the program. 3. Vendor Surveys: Completed for all Standard and Standard-Very Large NTGR projects that use vendors, as well as for Basic NTGR projects where customers indicated a high level of influence from

vendors on their decision. Vendors include contractors, design engineers, distributors and installers.

4. Utility and Program Staff Interviews: Conducted for Standard and Standard-Very Large projects to obtain more insight into the customer's decision to implement the project and the extent of the utility's and program's role in the decision, along with vendor contact information.

5. Other Information: Secondary research is performed for Standard-Very Large projects to obtain information from other sources.

	Basic NTGR	Standard NTGR	Standard -Very Large NTGR
Program File	√	√	√
Decision Maker Survey	Core	Core, Supplemental	Core, Supplemental
Vendor Survey	√[1]	√[1]	√
Utility & Program Staff Interviews	√[2]	√	√
Other Research Findings			√

[1]Only performed for sites that indicate a vendor influence score (N3d) greater than maximum of the other program element scores (N3b, N3c, N3g, N3h, N3l).

[2]Only performed for sites that have a utility account representative

Timing of measurement

As close to the project completion as possible

Free ridership methodology

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.

The battery also obtains a description of what the customer would have done in the absence of the program. If the implementation was not an early

	replacement action, information about the alternative measures is obtained, which is used to adjust the gross engineering savings estimate for partial free ridership. The survey is based on a core set of questions for the Basic NTGR, with additional questions for both Standard and Standard-Very Large NTGR projects. The additional questions probe for more detailed information regarding how much of an influence the program had on the decision relative to a customer's internal policies such as financial criteria, corporate policy, and standard practice for implementing such projects. This information is used to check for consistency in responses for a project. Standard-Very Large projects may be subjected to additional questions that arise from review of other information sources. An internally consistent "story" is then created from all of the available data to support the NTGR calculated for Standard-Very Large projects. Vendors are asked about 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.
Free ridership algorithm	The NTG Ratio is calculated as an average of three scores, representing either the highest response or the average of several responses to the relevant question(s) about the decision to install the measure(s). The three scores are: 1. Timing and Selection score: Reflects the influence from the most important factor in the customer's decision to implement the project at this time. The vendor survey results enter directly into this score for projects where a high level of vendor influence was reported. 2. Program Influence score: Reflects the relative importance of the influence of the program to non-program factors on the customer's decision to implement the project. The program influence score is divided by 2 if the decision maker reported learning about the program after the decision was made on the specific measures to implement. 3. No-Program score: Represents the likelihood of actions the customer may have taken or would take in the future in the absence of the project. 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. Each score is based on the maximum score, representing the most important factor in the customer's decision making. High scores that are inconsistent with other responses trigger consistency checks and can lead to follow-up questions to resolve the discrepancy.
Spillover methodology	None
Spillover algorithm	N/A

Peters, Jane S. and Marjorie McRae, Research Into Action, Inc. *Free-Ridership Measurement Is Out of Sync with Program Logic… or, We've Got the Structure Built, but What's Its Foundation?* ACEEE. 2008.

Year(s) implemented	2008
Sponsoring agency/utility	ACEEE
Key Issues Addressed	Paper discusses problems associated with free-ridership (FR) measurement in directing energy efficiency programs and policies as well as the measurement of free-riders through participant self-report. Suggest that program policy, design, and implementation decisions can be better informed through analysis of market changes than through FR measurement.
Background	Currently, most estimation methods rely on participant self-reports obtained during site visits or telephone surveys. Reasons why the authors feel there are problems with current approach: 1. Customer decision making model is not appropriate in current program designs that use multiple methods to influence behavior. 2. Psychological theories of how people explain their behaviors to themselves contraindicate the use of current FR methods (attribution theory, cognitive dissonance theory, and social desirability bias in survey responses suggest we are overestimating FR). 3. Intentions are modest predictors of behaviors going forward, and certainly should not be used as predictor of behaviors retrospectively. 4. Increased awareness of global warming will likely result in increases in self-report FR. In 1994, Windel and Peters suggested a decision framework would be useful in helping to investigate how participants decided to invest in a measure through a program. This approach has become somewhat standard practice (e.g., 2003 Massachusetts guidelines and guidelines developed they the California Public Utility Commission Energy Division and the Master Evaluation Contractor Team in 2007). Friedmann (2007) offers three examples of program designs that would likely accelerate the deployment of energy efficiency technologies and behaviors, but are risky for the utility that does not get cost recovery for savings attributed to FR: • Upstream/midstream market programs with incentives to manufacturers and distributors and retailers • Establishing long-term relationships at various levels of both the utility and the large office building manager/owners to enhance energy efficiency uptake. • Addressing data centers with a variety of measures in a holistic manner.
Conclusions	The more effective EE programs are in changing typical market behavior, the less accurate self-report FR estimation methods are. They propose FR be proxied by the market saturation rate for the efficiency action. Market saturation potentially underestimates FR as those who already want to take the action may seek out the program. Yet, the actions from this group are partially offset by spillover actions. The authors suggest policy makers set market targets (% of market share), and do market studies that track the progress toward increased market-share. This leaves open the possibility that incentives may be even more important for later adopters than for early ones.

Peters, Jane S., Ph.D. and Ryan E. Bliss, Research Into Action. *Fast Feedback Pilot: Existing Buildings and Production Efficiency Programs.* Prepared for Energy Trust of Oregon. March 10, 2010.

Year(s) implemented	July 2009, January 2010
Sponsoring agency/utility	EnergyTrust of Oregon
Sector(s)	Commercial and Industrial
Goal	Primary research questions were whether and how the various methods affect completion rates and responses to survey questions.
Key Issues Addressed	Timing—survey conducted close to completion of project on a rolling basis throughout the year Survey method—Paper, phone, web
Background	Previous participant surveys had asked respondent to recall details of program-supported projects that had been completed up to two years before. Pilot tested new approach to collecting rapid feedback from program participants and evaluate different survey methods (paper, telephone, and web) from program participants in Energy Trust's Existing Buildings (commercial) and Production Efficiency programs (industrial).
Free-ridership methodology	Each month, completed projects or projects near completion were assigned to one of three survey methods: paper/phone/web. PE projects requiring on-site verification were assigned to the paper survey; all unverified projects and EB site-verified projects were randomly assigned to phone or web survey method. Paper and phone resulted in higher completion rates than web. Believe Fast Track approach results in more accurate free ridership figures than estimates gathered from participants a year or more after project completion. *(NOTE however, that the authors comment that savings weighted free ridership was comparable to the last program evaluation for the Existing Buildings program and somewhat lower for Production Efficiency.)* Recommendations: • Continue Fast Feedback Approach with phone method as this provides more immediate feedback to program staff and simplifies data collection/management • Explore and test modifications to the current approach of how projects would have changed without program support (e.g., don't assume that continuing to use existing equipment implies no equipment upgrade). • Expand Fast Feedback approach to all major programs.
Eligible respondents	Commercial and Industrial program participants.
Types of measures	HVAC systems, compressed Air, VSDs, motors, pumps, lighting, refrigeration, insulation, renewable, and commercial clothes washers.
Free ridership questions for customers	The free ridership assessment was based on the methodology developed for the evaluation of the 2006-07 PE program and adapted for the evaluation of the 2006-07 EB program. The assessment consists of 3 elements: 1) how the project would have changed without program assistance; 2) the availability of funds to do the project without program assistance; and 3) the program's influence on the project. PROJECT CHANGE QUESTIONS--Respondents were asked how their project would have changed if they had not participated in the program. Responses were coded into one or more of the following categories: 1. cancelled the project altogether 2. postponed the project more than one year

	3. repaired existing equipment 4. kept using existing equipment 5. purchased less expensive equipment 6. installed less energy-efficient equipment (slightly, somewhat or significantly less efficient) 7. reduced the project size or scope 8. not changed the project at all 9. don't know AVAILABILITY OF PROJECT FUNDS QUESTION--Would firm have made available the funds needed to cover the entire project cost in hadn't received incentive (yes, no, don't know) PROGRAM INFLUENCE—Asked to rate the influence on how the program was done for several program elements—the incentive, the installation vendor or contractor, the program representative, and a technical study (if applicable). Five-point scale, with 1 being not at all influential and 5 being extremely influential.
Free ridership questions for vendors	None
Free ridership algorithm	Using the above questions, they calculated 2 scores: the *Project Change Score* (based on the project change questions and the availability of project funds question) and the *Program Influence Score*. Both scores ranged from 0 (no free ridership) to 50 (indicating high free ridership). *Project Change Score*—Score of 0=project would postpone more than one year, repair, or continue using existing equipment (without specifying other changes, such as reducing the project scope or using less expensive or less efficient equipment), or use significantly less efficient equipment. Score of 25=respondent would reduce the scope of the project or use less expensive or somewhat less efficient equipment, or indicated some change but did not indicate what would have been done. Score of 50=respondents would do the project exactly the same or would use slightly less efficient equipment. *Program Influence Score*—Score based on the highest rated influence from among the program incentive, the program representative, and the technical study if one was performed. Score of 0=high program influence; a score of 25=moderate program influence; and a score of 50=low program influence. These scores were then summed with a resulting sum score ranging in value from 0 to 100. The scores were interpreted as a percentage, indicating a range from no to total free-ridership. In cases where there was insufficient data to calculate one or both scores, they calculated 2 free ridership scores: 1) a low-scenario score, which assuming that the missing score was 0, and 2) the high scenario score, which assumed that the missing score was 50. To calculate a mean free ridership across all respondents, they also calculated a third free ridership score, which was the mid-point of the low scenario and high scenario scores. Free-ridership scores across all respondents were reported as means of the low-scenario, mid-point scenario or high-scenario scores. Respondents without missing data had the same scores included in the mean calculations for all three scenarios. Thus, the magnitude of the range between the low-scenario score and the high-scenario score was based on the number of respondents with missing data. Influence ratings were largely unrelated to survey method. Requires good coordination between program staff and evaluator to receive monthly project information.

	Increased expense as method requires monthly cleaning of project information for purposes of sampling, as well as increased costs for monthly survey management.
Spillover questions for customers	None
Spillover questions for vendors	None
Spillover algorithm	NA

Prahl, Ralph, Prahl & Associates, Goldberg, Miriam and Bobbi Tannenbaum, KEMA Inc, David Sumi and Bryan Ward, PA Consulting Group, and Tom Talerico and Rick Winch, Glacier Consulting Group. *Integrating Supply-Side Results with End-User Net-to-Gross Self Reports.* Memorandum prepared for the Public Service Commission of Wisconsin. July 2, 2008.

Year(s) implemented	2008
Sponsoring agency/utility	Public Service Commission (PSC) of Wisconsin
Key Issues Addressed	Establish a framework for the performance of supply-side (SS) research in order to supplement end-user self-reports (SR) done as the primary approach to NTG analysis Clarifies the approach when the Selection Framework has resulted in the use of end-user self-reports yet there is thought to be the potential for supply-side effects to call the veracity of the results into question
Background	The framework is intended to capture all potential situations on a continuum ranging from changes in the behavior of vendors directly participating in the program to changes in the behavior of all vendors in the market or markets targeted by the program (participating and non-participating). Criteria to decide whether to perform supplemental supply-side research includes: 1. The existence of a *plausible, credible, and specific* program theory predicting supply-side program effects, or some other sound logical or empirical basis for believing they are likely to exist. 2. Likelihood that the predicted effects can be meaningfully assessed through empirical research. 3. Likelihood that the needed research can be performed at reasonable cost, relative to the available budget and the likely impact. The burden of proof in point #1 above becomes stronger the more expensive the issue would be to research. Key provisos; 1. Not all SS effects imply immediate energy savings 2. It may not always be feasible to quantitatively integrate supply- and demand-side results for the following reasons: • Unavailability of sales data • Vendors and end-users may have perspectives that are difficult to reconcile quantitatively • May not be able to afford enough interviews with multiple categories of market actors to get a reliable picture of what is going on • Difficult to avoid double-counting when both supply- and demand-side savings are attributed to program. 3. When not possible to quantitatively integrate SS results with end-user SRs, qualitative conclusions will be drawn and presented to the PSC and the PSC can consider them in deciding how much credit to give a program 4. When SS and SR results are integrated, important to acknowledge the resulting uncertainties.
Integration Approaches	Different categories of integration of SS results with end-user SR:

	1. Making no changes to end-user self-report result • May be appropriate when supply-side results constitute leading indicators, or are not sufficient to conclude there are energy savings 2. Altering NTG approach for the next round • Use when there is evidence of supply-side effects, but not practical to incorporate the results into the current NTG analysis • Particularly appropriate when current energy savings are likely to be modest, but seem likely to increase over time • May be appropriate if supply-side research is exploratory, but yields results suggesting that more rigorous analysis may provide improved NTG estimates 3. Using supply-side results to refine end-user self-report battery • Example: program is found to have reduced incremental cost of energy efficient measures; ask end-users about WTP at higher price 4. Disentangling Supply- and Demand-Side Impacts • Use analytical techniques to eliminate or correct for overlap in the two net savings estimates, then adding them together. 5. Altering the self-report interpretative algorithm • Appropriate when supply-side results suggest significant changes in supply-side conditions of which end-users are unlikely to be aware. • Can override certain end-user SR responses based on SS results 6. Adjusting the self-report net-to-gross result • Prime example: supply-side research yields strong evidence of nonparticipant spillover effects, but does not call into question validity of end-user self-report responses 7. Not integrating the demand- and SS results • Evidence of SS effects, but none of the above approaches are appropriate

Rathbun, Pam, Carol Sabo, and Bryan Zent. PA Consulting Group. *Standardized Methods for Free-Ridership and Spillover Evaluation—Task 5 Final Report (Revised).* Prepared for the Massachusetts Utilities, June 13, 2003.

Year(s) implemented	2003
Sponsoring agency/utility	National Grid, NSTAR Electric, Northeast Utilities, Unitil, Cape Light Compact:
Sector(s)	Commercial, Industrial
Goal	Develop standardized methods to be used by all the sponsors to determine free-ridership and spillover factors for C&I programs.
Background	Previous studies in MA had used independent evaluation approaches, with varied survey instruments, analysis techniques and assumptions. This report represents a collaborative effort between sponsors and evaluators in developing standardized sampling techniques, data collection approaches, survey questions, survey instruments and an analysis methodology. The evaluation methodology is an assessment of the annual program impacts including disaggregated values for free-ridership and spillover.
Methodology	Recommendations for standardization were made in the following areas: **PRE-SURVEY PREPARTION** • Program application data should be maintained in an electronic database • Additional data should be collected at project closure to make it easier to identify the appropriate decision maker(s) and vendors. **SAMPLE DESIGN** • Samples should be designed to achieve a minimum of +/- 10% precision level at the 90% confidence level at the end-use measure category level. • A census of measures (i.e., all measures) should be included for end-use measure categories with less than 50 installations in order to achieve the minimum +/- 10% precision level. Precision levels worsen dramatically as the number of measures in the population decreases from 50 • For categories with more than 50 installations, a stratified sampling strategy should be implemented, sampling all top 10% savings sites plus a random sample of the remaining sites to achieve the minimum precision levels. • In all cases, a 60% response rate should be considered as a minimum goal when conducting the surveys. • Customers with multiple measure categories should only be asked about their decisions regarding two measures, with priority given to the more rare measures and/or measures with the largest savings. **SURVEY IMPLEMENTATION** • Surveys should be conducted within a year of participation • Surveys should be administered via telephone (though large, custom or industrial projects may require on-site surveys) by professional interviewers. • An advance letter on Sponsor letterhead explaining the study should be mailed to all sampled program participants prior to the survey.

	• The key decision-maker(s) must be identified when conducting the surveys. • When evaluating a customer's free-ridership rate, it may be necessary to interview the design professional or vendor involved in the project. **ANALYSIS:** • Free-ridership and spillover estimation should be conducted annually after the end of the program year. • Free-ridership and spillover estimation should be conducted at the specific end-use measure category level. • Completed surveys must be weighted to account for disproportional sampling probability and non-response so the results represent the population of measures. That is, statistical expansion methods appropriate to the sampling process are required. • With standardized methods to estimate free-ridership and spillover, Sponsors can also calculate net program savings in a consistent manner: *Net Savings = (Gross Savings) * (RR) * (1-FR+PS+NPS)* *=(Gross Savings) x (RR) x (1-FR+SO), where* *RR is the realization rate (evaluated/tracking), FR is the free-rider fraction, PS is the participant "like" spillover fraction, and NPS is the nonparticipant spillover fraction. SO (total spillover) is the sum of PS and NPS.* Variations on survey wording may be necessary to fit key decision-making groups that vary by program type.
Timing of measurement	Should be conducted annually after the end of the program year
Free ridership questions for customers	• Identification of key decision maker • Project and decision-making review questions: Intended as warm-up/context questions • Reminder of what incentive and services (e.g., technical assistance) were received • Timing: whether or not any measures would have been implemented within one year without incentive • Quantity: whether they would have purchased exact same quantity in absence of incentive • Program Efficiency: what percent of installed measures would have been of same efficiency without incentive • Cost: whether the company would have paid for the same installed measures in absence of program • Consistency checks for measures initially assigned free-ridership of 0% or 100% • Technical Assessment Study impact question • Past program participation impact questions
Free ridership questions for vendors	Confirmation that key decision maker was design professional. Rest of questions are parallel to the customer questions
Free ridership algorithm	Free-ridership (both pure and partial), using a customer survey: • Definition: customer who received an incentive who would have

	installed the same or smaller quantity on their own within one year if the program had not been offered • Calculation addresses the full range of total free-ridership (0% to 100%), based on the quantity and efficiency of any equipment that would have been installed outside the program • Other factors such as utility-sponsored technical assessments and the influence of past program participation are considered in the free-ridership rate calculation
Spillover questions for customers	Measured only like spillover--questions probe on recent purchases and the similarity to measures installed from the program, and the influence past installation through the program had on the decision.
Spillover questions for vendors	Four steps are used to determine nonparticipant "like" spillover: 1. For each design professional/vendor, the survey determines the percentage of all program-eligible equipment sold/installed outside the program in the Sponsor's service territory. 2. For each design professional/vendor, the survey determines whether the sale or installation of program-eligible equipment outside the program was due to the program (nonparticipant spillover). 3. For each design professional/vendor, savings associated with this "nonparticipant spillover" equipment are determined by examining the participant database and quantities installed. 4. Nonparticipant spillover savings are then extrapolated from the survey to the total program savings in the year.
Spillover algorithm	Participant like-measure spillover, using a customer survey: • Definition: customer who installed equipment through the program in the past year and then installed additional equipment of the same type ("like") • Calculation assigns weights based on the source of the spillover – experience with the equipment, participation in any past program, or recommendation from design professional Nonparticipant like-measure spillover, using a survey of participating design professionals and vendors: • Definition: energy efficiency measures installed by program nonparticipants due to program influence, based on responses from design professionals and vendors participating in the program • Calculations based on fraction of equipment receiving no program incentive and the program's influence on the design professional/vendor's decision to recommend equipment • The program tracking system database can be used to attach kWh savings estimates to nonparticipant spillover if the database contains this information

Ridge, Richard, Ken Keating, Lori Megdal, and Nick Hall. *Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approaches.* October 15, 2007.

Year(s)	2007
Sponsoring agency/utility	California Public Utilities Commission (CPUC)
Key Issues Addressed	Provides basic methodological guidelines that are considered best practice in the social science and engineering communities that evaluators should use to assess net impacts and spillover.
Background	The CPUC adopted the *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professions* (TecMarket Works, 2006) for the measurement and evaluation of energy efficiency programs. In certain situations, the Protocols allow for use of self-report approach (SRA) to estimate NTG for the basic, standard, and enhance levels of rigor. The protocols recognize that using both quantitative and qualitative data can be used to assess causality. However, the Protocols are silent regarding basic methodological guidelines that are considered best practice.
Free Ridership Methodology	The SRA deviates from the standard approach to assessing causality (experiment or quasi-experiment), which are not always desirable or possible. The SRA guidelines are as follows: 1. Timing of the interview—as soon after installation as possible. 2. Identifying the correct respondent(s)—is critical. For large C&I is complicated as different actors have different and complementary pieces of information about the decision making; decisions may be being made in regional or national headquarters; decision making may be done by commissions, committees, boards or councils; and there may be both a technical and a financial decision maker. 3. Set up Questions—Need to adequately establish the context and sequence of events that led to decisions. 4. Use of multiple questions—uses both quantitative and qualitative questions to measure a construct such as free ridership is preferable as reliability is increased. 5. Validity and reliability—should be assessed for each question used. Also the internal consistency of multiple-item scales should be tested. For large savings sites, multiple members of the evaluation team should review the results to ensure consistency. 6. Consistency checks—set up checks for inconsistencies and establish rules for handling inconsistent responses while the respondent is on the phone. 7. Make questions measure-specific 8. Partial free-ridership—should explore cases where participant would have installed something more efficient than program assumed baseline but not as efficient as program equipment. 9. Deferred free-ridership—need to measure program's impact on acceleration of installation. Should use more than one question and use preponderance of evidence approach. 10. Scoring algorithms—can impact results and must be documented and tested using sensitivity analyses. Preponderance of evidence approach is better than relying solely on an algorithm. 11. Handling "don't knows" and non-responses—need to determine in advance how these will be handled. Make a special effort to avoid

	don't know responses. 12. Weighting the NTGR—Need to take into account the size of the savings impacts at the customer or project level. 13. Ruling out rival hypotheses—need to ask open-ended questions regarding other possible reasons for installing the efficient equipment. 14. Precision of the estimated NTGR—need to report this but is complicated when have multiple sources of information or multiple respondents. In this case need to take into account the propagation of errors in the relative precision. 15. Pretesting the instrument—always pretest to reveal ambiguous wording, faulty skip patterns, leading questions, faulty consistency checks and incorrect sequencing of questions. 16. (Large Savers, complex decision making) Incorporation of additional quantitative and qualitative data in estimating the NTGR. • use multiple respondents • use other site- and market-level data • establish rules for data integration • analysis method (case studies are one method for assessing both quantitative and qualitative data, content analysis to identify coherent and important these and patterns in the data). Use of multiple evaluators to independently review the data. 17. Qualified interviewers—For complex situations, engineers familiar with the more complicated technologies should be trained to collect the data.

Ridge, Richard, Ridge & Associates, Phillipus Willems, PWP Inc, Jennifer Fagan, Itron, Inc., and Katherine Randazzo, KVD Research Consulting. *The Origins of the Misunderstood and Occasionally Maligned Self-Report Approach to Estimating the Net-To-Gross Ratio.* IEPEC. 2009.

Year(s) implemented	2009
Sponsoring agency/utility	Energy Program Evaluation Conference, Portland
Key Issues Addressed	While the SRA approach has been used for over 30 years, because it does not involve any formal comparison groups it has been criticized as inherently biased and unreliable. This paper discusses the role of the self-report approach (SRA) within the larger evaluation framework, improvements to the SRA over time in both its internal validity and reliability, and responds to more common criticisms of the California SRA.
Background	Which technique is used to estimate net energy and demand impacts depends on a number of factors, including time, money, data availability, and effect size. For example, the expected magnitude of savings for a given program might not warrant the investment in a billing analysis or discrete choice analysis. Or, key stakeholders might not want to wait for a billing analysis, which typically requires up to 12 months of post consumption data. And with a small signal to noise ratio, the sample sizes necessary for the required statistical power can be prohibitively expensive. In addition, there are situations in which the standard quantitative approaches involving comparison groups are not always possible. For example, in the industrial sector there are 3 barriers—low signal to noise ratio in a part./nonpart. billing analysis, industrial customers have participated in energy efficiency programs in prior years making it difficult to find true nonparticipants, and large industrial customers are each unique making it unlikely that one could find a matching nonpart. group. In new construction programs, many of the large res. and nonres. developers, architects and engineering firms have also been contaminated by prior participation in energy efficiency programs. The authors believe that discrete choice, difference-of-differences, and econometric modeling approaches have become over time increasingly unreliable and implausible in these cases and that SRA is the most appropriate for evaluating this complex and diverse program and market. In 1993, California formally recognized that methods involving comparison groups were not always feasible. The1993 California Protocols allowed the SRA as one way to estimate the NTGR. The 2005 Protocols expanded on the SRA approach and explicitly required triangulation for programs assigned the enhanced level of evaluation rigor understanding that there is error associated with any single method. Triangulation uses a variety of research methods and data sources to reduce the risk of systematic biases.
Free Ridership Methodologies	The establishment of a causal connection between the program and customer behavior is at the core of arguments for and against research methods aimed at establishing program impacts. One strategy is to develop and assess rival hypotheses to guard against threats to validity. Sound program theories and logic models can provide valuable assistance in identifying the plausible rival hypotheses. Approaches to demonstrating causality using non-experimental designs could include case studies or SRA. In 2007, the Energy Division published the *Guidelines for Estimating Net-to-Gross Ratios Using the Self-Report Approaches* which contained 17 recommendations for further improving the validity and reliability of the CA-SRA. In 2007, the CPUC also formed two groups (the Residential and Non-Residential NTGR Working Groups) of nationally recognized evaluators to consolidate the lessons learned over the last 15 years in order to make further improvements in the CA-SRA. The CA-SRA involves asking one or more key decision makers a series of closed

	and open-ended questions about their motivations for installing the program-eligible equipment, about what they would have done in the absence of the program, as well as questions that attempt to rule out rival explanations for the installation. In the simplest case (e.g., residential customers), the CA-SRA is based mainly on quantitative data. In more complex cases in nonresidential programs, the CA-SRA is strengthened by including additional quantitative and qualitative data (e.g., in-depth open-ended interviews, direct observation, review of customer and program records).
Response to criticisms of SRA	The authors note a number of criticisms of the SRA and provide their responses to those criticisms. 1. Legitimacy—CA-SRA is a legitimate social science tool for establishing causality. 2. Turbulent environment—In any evaluation, as the number of alternative hypotheses grows, the task of teasing out the single intervention effect becomes more challenging. This is the case regardless of what method is used. 3. Nonlinear approach—the CA-SRA recognizes that the route is nonlinear by attempting to identify other parties most important in a customer's decision to participate and uncover various ways the program might have influenced these market factors. 4. Recall—Interviews should be conducted with the decision maker as soon after the installation of equipment as possible. 5. Subjective—the CA-SRA collects a variety of qualitative and quantitative evidence so it is not merely subjective. 6. Treating ordinal data as interval—there is strong support in the social science literature that treating ordinal scales as interval data yields results that are both meaningful and useful to decision makers, and there is no reason to think that this measurement is not randomly distributed. 7. The meaning and calculation of NTGR—algorithms and weights must be developed by experienced professionals who understand that these have to be transparent, plausible and defensible and must be subjected to sensitivity analysis. 8. Socially desirable responses—Methods have been developed to address this potential source of bias. These methods have been incorporated into the CA-SRA. 9. Stated intentions—The CA-SRA collects a variety of information to measure the counterfactual, get at the main reasons for installing the efficient equipment, and establish the temporal precedence of the program. For more complex projects additional information is gathered from vendors and file review.
Conclusions	The paper concludes by saying it doesn't make sense to compare all SRA approaches equally as some conform to best practices (CA-SRA) and others don't. For projects with substantial savings that have been assigned the enhanced level of rigor, the 2005 Protocols require that two or more approaches of the available three (discrete choice with a comparison group, billing analysis with a comparison group, and the CA-SRS) must be used. For programs that have been assigned the standard or basic level or rigor and for which methods involving comparison groups are impossible, the CA-SRA can provide sufficiently rigorous estimates of NTGR. Any set of rewards and penalties should never require a level of accuracy that exceeds the ability of any evaluators to provide. This is unreasonable burden on evaluators and results in a continuing, contentious and unproductive relationship

	among implementers and regulators.

Saxonis, William P. New York State Department of Public Service. *Free-Ridership and Spillover:*
A Regulatory Dilemma. IEPEC. 2007.

Year(s)	2007
Sponsoring agency/utility	Energy Program Evaluation Conference, Chicago
Key Issues Addressed	The paper examines free rider and spillover results from energy efficiency programs administered by NYSERDA, which found that free rider rates for C&I programs ranged from 10-67% and spillover rates ranged from 19-168%. For residential programs, the free ridership ranged from 2-28 percent and spillover from 5-48 percent. Paper looks at FR and SO measurement in a historical context, compares NYSERDA results to other states, and concludes with practical recommendations.
Background	In NY, NTG=(1-FR)+SO Regulators need reliable estimates for 3 reasons: 1. Protect ratepayers economic interests 2. A secure supply of electricity. 3. Environmental Spillover measurement trails free ridership measurement in the level of research attention and the level of confidence in the reliability of the results.
Methodology	Due to size and scope of the NYSERDA program portfolio there may be some variations in the following general FR approach used by NYSERDA: • Directly ask participants if they would have implemented the same energy efficiency measure without program assistance • Ask quantitative and open-end questions regarding program influence • Score open end responses using an established formula to capture the degree of FR based on factors such as the timing of installation, quantity, and efficiency. The spillover approach: • Multi-question survey approach similar to FR methodology with participating customers/vendors. • For some programs, nonparticipants were surveyed to determine any influence of the program on their energy efficiency related behavior. For some evaluations, used an “integrated data collection process” to gain participant feedback in near real time to supplement retrospective survey efforts. Participants were asked to complete an abbreviated survey containing questions related to program attribution soon after their participation. This approach is useful for identifying trends and confirming FR and SO values in between major evaluation cycles.
Analysis	Despite the high FR and SO rates in NYSERDA programs (especially compared to other programs in the region), the impact on NTGR is virtually non-existent.
Conclusions	1. Improve data reliability—has been little research to quantify FR and SO results using multiple approaches in the same study. This would help increase confidence in the data if they produce similar results.

	Need to increase methods to triangulate the data, increase the precision and CI of surveys, employ more long term and comparative analysis (especially for SO), conduct studies that compare adoption of ee products in regions with and without intervention programs, and develop more probing questions that go beyond questions related to specific actions. 2. Leverage FR/SO data to maximize value—need to understand the change in FR/SO levels as economic conditions and markets evolve so that programs meet today's needs. Monitor program application rates, process evaluations, product baselines, etc. Link FR/SO data with results from questions on demographics, attitudes toward environment and energy efficiency, reasons for program participation, shopping preferences and status of economy and conduct longitudinal studies to see how rates change over time and under what conditions. 3. Increase collaboration—need to look at attribution in both regional and national forums. By doing this on a group basis, could defray costs.

Stoops, John, KEMA, Inc., The Cadmus Group, Inc., Itron, Inc., and Nexus Market Research, Inc. *Non-Residential New Construction (NRNC) Programs Impact Evaluation.* California Public Utilities Commission Energy Division. February 08, 2010.

Year(s) implemented	2006-2008 Program Years
Sponsoring agency/utility	CPUC
Sector	Commercial and industrial new construction (plus agriculture new construction for PG&E)
Goal	Savings by Design provides design assistance and financial incentives to improve the energy efficiency of commercial new construction
Timing of Measurement	Projects completed in 2006-2008 PY, but may have been started in earlier PYs.
Eligible respondents	Program participants – decision makers including building owners/managers, architects and engineers
Type of measures	System Shell, System Lighting, System HVAC + Motors, System Refrigeration, Whole building
Free ridership questions for customers	Most SBD have multiple measures, for which the levels of free-ridership may vary across the measures. Consequently, the survey questions were asked for every incented measure in the tracking database (systems approach) or identified in the project file (whole building approach). The decision maker survey included the following questions as the NTG battery: • On a scale from 0 to 10, where 0 means not influential whatsoever and 10 means extremely influential, How influential was Savings by Design, including the incentives, design assistance, design analysis and interactions with SBD representatives and consultants in the implementation of <measure description>? • How did Savings by Design influence the implementation of <measure> (choose all that apply)? • On a scale from 0 to 10, where 0 means that this measure would have been installed exactly the same regardless of interaction with Savings By Design regarding this project and 10 means that the measure would definitely not have been installed without SBD influence and interaction, what is the likelihood that this measure would not have been installed with SBD interaction? Why?
Free ridership questions for vendors	None
Free ridership algorithm	Since most projects involved multiple measures with potentially a range of free-ridership values, the estimation of the NTG ratios incorporated the responses for all measures. The responses to the “influence of the program” and “in the absence of the program” questions were used to generate measures’ net-savings scores which were used to estimate measures’ NTG ratio. The “how influential” question response was multiplied by 0.1 to assign points for that response. The “how did SBD influence the implementation” question was assigned a score of 0 to 2 points, based on the response given. The “in the absence of the program” question was assigned points by multiplying the answer by 0.3. The cumulative score for each measure was compared to the max value of 6 to determine the degree of free-ridership. A score of 6 indicates that the measure was completely influenced by the program, and a score of zero

	indicates the measure would have been installed without the influence of the program. The responses from the decision maker surveys were reviewed along with the program file to assess for consistency.
Spillover questions for customers	None
Spillover questions for vendors	None
Spillover algorithm	N/A

Winch, Rick and Tom Talerico, Glacier Consulting Group, Bobbi Tannenbaum, KEMA Inc., Pam Rathbun, PA Consulting Group, and Ralph Prahl, Prahl & Associates. *Framework for Self-Report Net-to-Gross (Attribution) Questions.* July 2, 2008.

Year(s) implemented	2008
Sponsoring agency/utility	Public Service Commission (PSC) of Wisconsin
Goal	Develop a framework to guide the revision of existing survey instruments used for determining attribution.
Key Issues Addressed	Revise existing surveys to improve consistency across Focus program areas and provide transparency for the approaches used.
Background	A working group of the FOE Evaluation team reviewed the existing self-report attribution batteries. The reviewers identified four areas of information that ideally would be collected as part of a self-report battery: 1. Context: recollection of past events, the sequence of these events and the how these events affected the participation process. 2. Decision-making: Having participants discuss their Focus project-related decision making; identification of factors that contributed to the process, and what decision-makers were involved. 3. Direct Attribution: Assess the impact of the program on the timing, efficiency level and quantity of technology installed. 4. Consistency Checks: Identified both during the interview and in the analysis stage, level of effort to resolve varies by importance of case. a. CATI survey implementation: changes to initial calculation of attribution are made by the analyst or project manager in the analysis phase b. In-depth project review by senior staff: Interviewer should be provided guidance on where to anticipate inconsistencies and relied upon to make a judgment of the degree of influence the program had on the customer responses.
Free ridership methodology	Specific questions were not developed, but types of information or issues to be considered are provided: ***Context*** *Information classified as "Key Information"* 1. Confirm or determine whether the project involves new construction, building expansion, replacement of existing equipment, or modification to existing equipment. 2. Confirm type of equipment installed, date, reward amount, and other items deemed relevant. 3. Confirm evaluator's information regarding key services, rewards and assistance provided by Focus as well as the type and amount of vendor/implementer involvement. 4. Determine when and how respondent first heard about the services/rewards/assistance available through Focus. • Explore possibility that new equipment was already installed before hearing about the services/rewards/assistance available from Focus. • Explore any plan(s) to purchase or install equipment before learning about the services/rewards/assistance available through Focus. a. Understand existing plans.

b. Understand point in planning process that respondent/organization (1) became aware of Focus and (2) begin discussing plans with Focus representative(s).

c. Understand qualitatively the impact/changes necessitated by Focus involvement.

5. Discuss the working condition of replaced equipment (Probe: planned replacement/upgrade, failure, estimated remaining useful life, repair history)

Information classified as "Supporting Information"

1. Explore what first made respondent (organization) start thinking about installing/replacing equipment at (home/this facility).
2. Age of equipment that was replaced.
3. Explore previous Focus participation.

Decision-Making

Information classified as "Key Information"

1. Organizational policies that specify factors considered when purchasing new (replacing old) equipment/ (Probe: payback, return on investment, guidelines on efficiency levels)
2. Major obstacles/barriers faced when seeking approval for project. (Probe: budget, time constraints, other priorities, disruption of production, etc.)
3. Role of contractor(s)/vendor(s) in project.

- Making respondent aware of Focus (or vice versa).
- Decision to participate.
- Recommendation to install certain type/energy efficiency level of equipment.
- Influence of contractor/vendor involvement on decision to install equipment at this time.(If not available from database)

4. Explore the percentage of the total costs—that is, all financial assistance plus the costs not covered by financial assistance—of installing improvements that were covered by Focus.

Information classified as "Supporting Information"

1. Budgeting process for new/replacement equipment. (Probe: size projects budgeted for, budget planning cycle/length)
2. Who within organization is responsible for recommending the purchase of new/replacement equipment.
3. Who within organization is responsible for approving the purchase of new/replacement equipment.

Attribution (Timing, Efficiency, Quantity)

Timing

Information classified as "Key Information":

[Note: Remind respondent of ALL services/rewards/assistance from Focus—spanning from a facility assessment to educational materials to rebates/rewards].

1. Explore whether or not, in absence of any assistance from Focus, respondent would have replaced equipment (purchased new equipment) at the same time.

- If no, determine when it would have taken place. (If they cannot give

number of years/months, then probe with mutually exclusive response categories.)

This is a critical point for a consistency check with responses in the context and decision making sections.

Efficiency

Information classified as "Key Information"

(If operate multiple facilities)

1. Explore whether or not (before installing this equipment) the organization had installed equipment of the same energy efficiency level at this or another facility without receiving services/rewards/assistance like those from Focus.
2. Explore whether or not respondent, when considering the purchase of this equipment, was aware of ***a range of*** efficiency levels that could have been chosen. Explore the range.

- Explore respondent understanding of efficiency levels available prior to Focus involvement.
- Explore when/how respondent first became aware of the energy-efficiency options available.
- Explore role of Focus (Focus representatives) in helping respondent understand the range of efficiencies available.
- If respondent has difficult time answering previous three questions, then: Explore whether or not all available efficiency options presented to respondent qualified for a reward/incentive through Focus (i.e., did respondent have both rebated and non-rebated option?)
- Without the services/reward/assistance from Focus, ask if respondent (organization) would have installed less efficient equipment. (Probe: If efficiency range available, ask more pointed question about what the efficiency level would have been).

Information classified as "Supporting Information"

1. Explore respondent's awareness of how the efficiency level of the old equipment compares to the efficiency level of the equipment that replaced it. (If new equipment must meet a minimum governmental standard, probe to see if respondent understands what that standard is.)

Quantity

(ASK QUANTITY MODULE ONLY OF RELEVANT MEASURES WHERE THERE IS A VARYING LEVEL OF QUANTITIES—E.G., LIGHTING, MOTORS, VSD)

Information classified as "Key Information"

1. Determine whether or not respondent (organization) would have installed the same quantity of equipment, fewer, or more at that time without the program.
2. If more or less, explore what percentage of (measure) would have been installed without the program assistance.

Consistency Checks

(Questions that can be used to check consistency were included in the Context, Decision-Making and Direct Attribution sections. The questions below work best at the end of the attribution section of a survey.)

Information classified as "Key Information"

1. Explore importance of services/rewards/assistance received from

	Focus on decision to install (measure)? 2. Ask respondent to describe, in their own words, what they (their organization) would have done if they had not participated in Focus (i.e., not received the Focus-related services/awards/assistance that they did).
Free ridership algorithm	Responses to Direct Attribution are compared to Context and Decision-making responses to identify inconsistencies. Level of effort for processing a single case during the analysis phase depends on the importance of the case to the overall attribution rate. Analysis of self-reported battery varies by: 1. Complexity of project 2. Number of decision-makers 3. Role of suppliers 4. Evaluation budget For example, a program may have a large number of participants, relatively small or simple projects, single decision-makers and little involvement of suppliers. The analysis may be straight-forward, using direct attribution questions along with a few consistency checks. The data may be collected through CATI surveys. Conversely, a program or project may have a small number of large, complex projects, multiple decision makers and multiple suppliers. The analysis would likely involve reviewing information from multiple sources to form an overall picture of the program or project. This information would be considered to make an informed decision as to the likely action (or lack thereof) that would have taken place in the absence of the program.

APPENDIX C: BIBLIOGRAPHY OF REFERENCES

Cook, Gay, Summit Blue Canada Inc. *Attribution Methodology Wars: Self-Report Methods versus Statistical Number Crunching—Which Should Win?* Paper delivered at ACEEE. 2008.

Cooney, Kevin, Beth Baker, Timea Zentai, and Adam Knickelbein, Summit Blue Consulting, LLC, *Gas Furnace Market Transformation Model Development and Market Research.* Submitted to Energy Trust of Oregon. August 5, 2009. Presented by Fred Gordon at AESP Brownbag. 2010.

Dohrmann, Donald, ADM Associates, Inc. John Peterson, Athens Research. Steve Westberg, Hiner and Partners. John Reed, Innovologie. *Evaluation Study of the 2004-05 Statewide Residential Appliance Recycling Program.* April 2008.

Efficiency Vermont. *Technical Reference User Manual (TRM), No. 2005-37, Measure Savings Algorithms and Cost Assumptions Through Portfolio 37.* Sent to Vermont Department of Public Service. November 29, 2005.

Erickson, Jeff and Mary Klos, Summit Blue Consulting and Valy Goepfrich, WPPI Energy. *Free-Ridership: Arbitrary Algorithms vs. Consistent Calculations.* IEPEC. 2009.

Fagan, Jennifer, Mike Messenger, and Mike Rufo, Itron, Inc. and Peter Lai, CPUC Energy Division. *A Meta-Analysis of Net-to-Gross Estimates in California.* AESP. 2009.

Goldberg, Mimi L., J. Ryan Barry, Erika Morgan, Ben Jones, Joshua Horton, and Nicole Buccitelli, KEMA, Inc. *Business Programs: Additional Looks at Attribution.* Prepared for the Public Service Commission of Wisconsin. February 26, 2010.

Goldberg, Mimi, KEMA Inc, Rick Winch and Tom Talerico, Glacier Consulting Group, Ralph Prahl, Prahl & Associates, and Bryan Ward, PA Consulting Group. *Treatment of Accelerated Savings.* Memorandum prepared for the Public Service Commission of Wisconsin. July 2, 2008.

Goldberg, Miriam L., J. Ryan Barry, Tammy Kuiken, Ben Jones, Paulo Tanimoto, Nicole Buccitelli, Colin Rickert, and Darcy DeAngelo-Woolsey; KEMA, Inc., *Business Programs: Acceleration Treatment and Life Cycle Net Savings.* Submitted to the Public Service Commission of Wisconsin. March 10, 2010.

Goldberg, Miriam, KEMA Inc., Oscar Bloch, Wisconsin Department of Administration, Ralph Prahl, Prahl & Associates, David Sumi and Bryan Ward, PA Consulting Group, and Rick Winch and Tom Talerico, Glacier Consulting Group. *Net-to-Gross Method Selection Framework for Evaluating Focus on Energy Programs.* Prepared for Public Service Commission of Wisconsin. March 16, 2006

Itron, Inc. and KEMA. *2004/2005 Statewide Express Efficiency and Upstream HVAC Program Impact Evaluation*. December 31, 2008.

Keating, Kenneth M., PhD. *Free-Ridership Borscht: Don't Salt the Soup.* IEPEC. 2009.

KEMA, Inc. *Evaluation of the 2004-2005 Statewide Multifamily Rebate Program–Volume I.* March 16, 2007.

Keneipp, Marshall, Floyd Keneipp, and Jeff Erickson, Summit Blue Consulting, LLC and Bill Norton, Opinion Dynamics Corp. *APS Measurement, Evaluation, & Research (MER) Report, Consumer Products Program (CPP).* APS. September 30, 2008.

Klos, Mary and Joan Huston, Summit Blue Consulting, LLC. *Impact Evaluation of 2007 CFL Buy-Down Pilot.* Prepared for Progress Energy—Carolinas. May 20, 2008

Lee, Allen, Cadmus, and KEMA, Inc. *Codes & Standards (C&S) Programs Impact Evaluation.* April 9, 2010.

Lori Megdal, Rick Ridge, Pam Rathbun, PA Consulting. Mimi Goldberg, KEMA. Ben Bronfman Quantec. Stuart Schare, Summit Blue; Nick Hall, Ken Keating. *Joint Simple Net of Free-Ridership and Participant Spillover SRA Battery.* April 19, 2008.

Megdal, Lori, Megdal & Associates, LLC, Yogesh Patil, Energy & Resource Solutions, Inc., Cherie Gregoire and Jennifer Meissner, New York State Energy Research and Development Authority,

and Kathryn Parlin, West Hill Energy & Computing, Inc. *Feasting at the Ultimate Enhanced Free-Ridership Salad Bar*. IEPEC. 2009

National Action Plan for Energy Efficiency (2007). *Model Energy Efficiency Program Impact Evaluation Guide.* Prepared by Diane Munns and Jim Rogers. <www.epa.gov/eeactionplan>

Nonresidential Net-To-Gross Ratio Working Group. *Methodological Framework for Using the Self-Report Approach to Estimating Net-to-Gross Ratios for Nonresidential Customers.* May 8, 2009.

Peters, Jane S. and Marjorie McRae, Research Into Action, Inc. *Free-Ridership Measurement Is Out of Sync with Program Logic… or, We've Got the Structure Built, but What's Its Foundation?* ACEEE. 2008.

Peters, Jane S., Ph.D. and Ryan E. Bliss, Research Into Action. *Fast Feedback Pilot: Existing Buildings and Production Efficiency Programs.* Prepared for Energy Trust of Oregon. March 10, 2010.

Prahl, Ralph, Prahl & Associates, Goldberg, Miriam and Bobbi Tannenbaum, KEMA Inc, David Sumi and Bryan Ward, PA Consulting Group, and Tom Talerico and Rick Winch, Glacier Consulting Group. *Integrating Supply-Side Results with End-User Net-to-Gross Self Reports.* Memorandum prepared for the Public Service Commission of Wisconsin. July 2, 2008.

Rathbun, Pam, Carol Sabo, and Bryan Zent. PA Consulting Group. *Standardized Methods for Free-Ridership and Spillover Evaluation—Task 5 Final Report (Revised).* Prepared for the Massachusetts Utilities, June 13, 2003.

Residential/Small Commercial Joint Simple Net-to-Gross (Self-Report) Committee, *Response to Overarching Comments Regarding the Use of Self-Reported Net-to-Gross (NTG) and the Residential and Small Commercial Self-Report Approach NTG Method*, January 28, 2010.

Ridge, Richard, Ken Keating, Lori Megdal, and Nick Hall. *Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approaches.* October 15, 2007.

Ridge, Richard, Ridge & Associates, Philippus Willems, PWP Inc, and Jennifer Fagan, Itron, Inc. *Self-Report Methods for Estimating Net-to-Gross Ratios in California: Honest!*. AESP. 2009.

Ridge, Richard, Ridge & Associates, Phillipus Willems, PWP Inc, Jennifer Fagan, Itron, Inc., and Katherine Randazzo, KVD Research Consulting. *The Origins of the Misunderstood and Occasionally Maligned Self-Report Approach to Estimating the Net-To-Gross Ratio.* IEPEC. 2009.

Saxonis, William P. New York State Department of Public Service. *Free-Ridership and Spillover: A Regulatory Dilemma*. IEPEC. 2007.

Skumaz, Lisa, SERA, Sami Khawaja and Jane Colby, Cadmus Group. *Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence of Energy Efficiency Behavior.* November 2009.

Stoops, John, KEMA, Inc., The Cadmus Group, Inc., Itron, Inc., and Nexus Market Research, Inc. *Non-Residential New Construction (NRNC) Programs Impact Evaluation.* California Public Utilities Commission Energy Division. February 08, 2010.

TecMarket Works, et al. *The California Evaluation Framework*. June 2004.

Titus, Elizabeth and Michals, Julie, Northeast Energy Efficiency Partnerships. *Debating Net Versus Gross Impacts in the Northeast: Policy and Program Perspectives.* AESP. 2008.

Winch, Rick and Tom Talerico, Glacier Consulting Group, Bobbi Tannenbaum, KEMA Inc., Pam Rathbun, PA Consulting Group, and Ralph Prahl, Prahl & Associates. *Framework for Self-Report Net-to-Gross (Attribution) Questions.* July 2, 2008.

Wright, Roger and Stoops, John, RLW Analytics. *An Evaluation of the 2004-2005 Savings By Design Program.* October 2008.

APPENDIX D: COMPARISON OF MEDIAN AND MEAN SUBSTITUTIONS

Table D-1. Different Free-ridership Scoring Scenarios

		Free-ridership with median substitution	Free-ridership with mean substitution	N	Free-ridership without median substitution	N
Energy Initiative	Lighting	17.4%	13.5%	39	4.4%	32
	Compressed Air	26.7%	26.7%	6	26.7%	6
Small Business Services	Lighting	3.6%	3.6%	36	3.6%	36

Table D-2. Free-ridership by Corporate Policy Regarding New Equipment

R3 - Company has energy efficiency policy when purchasing new equipment	Free-ridership Rate with Medians	Free-ridership Rate with Means	Number of Participants
Yes	32%	25%	19
No	10%	9%	60
Don't know	0	0%	2
Total	**16%**	**13%**	**81**

Table D-3. Free-ridership by Type of Corporate Policy

R4 - Type of energy efficiency policy at company	Free-ridership Rate with Medians	Free-ridership Rate with Means	Number of Participants
Purchase energy efficient measures regardless of cost	11%	11%	1
Purchase energy efficient measures if it meets payback or return on investment criteria	7%	7%	16
"All of the above"	30%	30%	1
Corporate policy requires annual energy reduction	100%	71%	1
Total	**32%**	**25%**	**19**

Table D-4. Free-ridership by Important Factors in New Purchases

R4B - Most important factor in new equipment purchases	Free-ridership Rate with Medians	Free-ridership Rate with Means	Number of Participants
Equipment performance	37%	28%	11
The equipment's energy consumption	20%	20%	26
Contractor, design professional, engineer's recommendations	10%	10%	2
Lifetime cost of the equipment	12%	9%	4
First (or up-front) cost of the equipment	11%	8%	17
Payback	7%	7%	10
Something else	1%	1%	5
Purchasing 'green' equipment	0%	0%	1
Program administrator recommendations	0%	0%	1
Meeting code	0%	0%	1
Don't know	0%	0%	3
Total	**16%**	**13%**	**81**

Table D-5. Free-ridership by Factors in Implementation Decision

R5 - Factors in decision to implement rebated projects	Free-ridership Rate with Medians	Free-ridership Rate with Means	Number of Participants
Factors in decision included program influence	9%	8%	31
Factors in decision outside of program influence	18%	14%	50
Total	**16%**	**13%**	**81**

Table D-6. Free-ridership Correlations by Participant Characteristics

Participant Characteristics	Pearson Correlation with Free-ridership Rate with Medians (n = 81)	Pearson Correlation with Free-ridership Rate with Means (n = 81)
Cost primary concern when making new purchases	-0.123	-0.133
Equipment primary concern when making new purchases	0.193	0.194
Only equipment factored in decision to implement	0.047	0.022
Corporate policy on energy efficiency	0.000	-.002
Pre-existing budget before speaking with the program	0.325	0.300

APPENDIX E: PARTICIPANT AND VENDOR SURVEYS FOR FULL-SCALE STUDY

2010 Participant Free-Ridership and Spillover Survey[36]

Variable List

<INTERVIEWER> = Interviewer Name
<CONTACT> = Customer Contact Name
<PROGRAM> = Program Name
<PA> = Program Administrator
<PA CONTACT INFORMATION> = PA Contact Name and Phone Number.
<CUST> = Customer/Facility Name
<DATE> = Date of participation
<YEAR> = Year of participation
<FUEL> = electric or natural gas
<ADDR> = Service address where measure was installed
<MEASCAT1, MEASCAT2> = End-use Category (i.e. lighting)
<QTY1, QTY2>
0 = quantity is not applicable for this measure category (measure count = 1 or quantity is not relevant as in delamping, recycling)
1 = quantity greater than 1
<EFF1, EFF2>
0 = efficiency is not applicable for this measure category (e.g., insulation, VFD, delamping, recycling, occupancy sensors)
1 = efficiency is applicable
<EQUIP1, EQUIP2> = 0 if installed measure is not equipment that is operational (e.g., insulation), 1=if installed measure is operational
<**MEAS1a-MEAS1h**>, <**MEAS2a-MEAS2h**> = detailed measure descriptions
<STUDY> = Technical Assessment Study, Technical Feasibility Study, Audit
<TA%> = Percent of study costs paid by PA
<TACOST> = Total cost paid for study
<TOINC> = Total incentive
<INC1, INC2> = PA incentive for specific measure categories
<TOTCOST> = Total project cost (customer cost+PA cost) for an account (by program)
<ALL ASSISTANCE> = Description of all technical assistance, financing, and rebates for measures installed through program
<FINANCE> = project received interest-free financing

NOTE:
For all questions, "DON'T KNOW" and "REFUSED" will be coded if offered as a response. Interviewers will probe as needed to minimize the amount of missing data.

[36] This instrument incorporates revisions made based on the pretest results.

For any case where the interview terminates early, respondent doesn't recall measures, measures are not installed, or the contact no longer work at the company and we cannot locate a knowledgeable respondent, the case will be pulled and sent to the PA for review.

Introduction

Hello, my name is <INTERVIEWER>, and I'm calling on behalf of <PA> regarding your firm's participation in their <PROGRAM>. May I please speak with <CONTACT>?

1 Yes
2 No [ATTEMPT TO CONVERT. MENTION ADVANCE LETTER THEY SHOULD HAVE RECEIVED REGARDING THE CALL.]

I1 Are you the person who was most involved in making the decision to get <ALL ASSISTANCE> through the <PROGRAM> in <DATE> at <ADDR> in <CITY>?

1 Yes [SKIP TO I2]
2 No [SKIP TO I1A]
D (DK) [PROBE TO IDENTIFY SOMEONE RESPONSIBLE FOR MAKING DECISIONS ABOUT ENERGY USING EQUIPMENT AT THAT FACILITY; IF DK, THANK AND TERMINATE]
R (REFUSED) [THANK AND TERMINATE]

I1a. Who was primarily responsible for making the decision to get <ALL ASSISTANCE> through the program?

[RECORD NAME AND DISPOSITION]

1 Transfers you
2 Can only give contact information [RECORD CONTACT INFO; THANK AND TERMINATE]
D (DK) [THANK AND TERMINATE]
R (REFUSED) [THANK AND TERMINATE]

I2. Are you employed by <CUST> or are you a contractor who provides design and/or installation services for <CUST>?
(INTERVIEWER NOTE: CODE UNPAID MEMBERS OF AN ADVISORY BOARD OR COMMITTEE AS EMPLOYEES)

1 Work directly for company/Employee/Volunteer
2 Vendor/Contractor [TERMINATE and USE VENDOR SURVEY]

INTRO1.

I'm with Tetra Tech, an independent research firm. On behalf of <PA>, we are following up with customers who participated in the <PROGRAM> in <YEAR> to learn about their experiences. You or someone at your facility may have received a letter from <PA> letting you know to expect this call. I'm not selling anything, I'd just like to ask about the energy efficiency project you implemented through this program at <ADDR>. Your individual responses will be kept confidential by Tetra Tech and <PA> This should take about 15 minutes.

Before we start, I would like to inform you that for quality control purposes, this call will be recorded and monitored.

READ FOLLOWING ONLY AS NEEDED:
(Sales concern: I am not selling anything; I simply want to understand what factors were important to your company when deciding to implement this new energy efficiency project and receive an incentive through this program. Your responses will be kept confidential by our firm and <PA>. If you would like to talk with someone from <PA>, you can call <PA CONTACT INFORMATION>.)

(Who is doing this study: <PA> has hired our firm to evaluate the program. As part of the evaluation, we're talking with customers that participated in the program to better understand their experiences with the program.)

(Why are you conducting this study: Studies like this help <PA> better understand customers' need for and interest in energy efficiency programs and services, and to improve the effectiveness of their programs.)

(Timing: This survey should take about 15 minutes of your time. Is this a good time for us to speak with you? IF NOT, SET UP CALL BACK APPOINTMENT OR OFFER TO LET THEM CALL US BACK AT 1-800-454-5070.)

Decision Making

INTRO2.

In the remainder of this interview, I'd like to focus on the <MEASCAT1, MEASCAT2> you implemented through the <PROGRAM>.

REPEAT R1A THROUGH R1D FOR MEASCAT1 AND MEASCAT2.

R1a. According to our records, the [EFFICIENCY IS APPLICABLE (IF EFF1, EFF2 = 1): energy efficient] <MEASCAT1, MEASCAT2> project you implemented through the program included <MEAS1a-MEAS1h, MEAS2a-MEAS2h>.

Were you involved in the decision-making process when the [EFFICIENCY IS APPLICABLE (IF EFF1, EFF2 = 1): energy efficient] <MEASCAT1, MEASCAT2> was being considered for this facility?

1 Yes
2 No
D (DK)
R (REFUSED)

R1b. Aside from yourself, who else within your company or outside your company was involved in the decision of whether or not to purchase the [EFFICIENCY IS APPLICABLE (IF EFF1, EFF2 = 1): energy efficient] <MEASCAT1, MEASCAT2> through the <PROGRAM>?

(PROBE: IF MORE THAN ONE DECISION MAKER, ASK R WHO WAS RESPONSIBLE FOR MAKING THE ULTIMATE DECISION)

1 No one else
2 (SPECIFY):

Name	Title	Phone number	Probe for role:

R1c. Is this <MEASCAT1, MEASCAT2> equipment still at least partially installed [IF INSTALLED MEASURE IS OPERATIONAL; (IF EQUIP1, EQUIP2=1): and operating] at this facility?

1 Yes [SKIP TO NEXT MEASURE]
2 No
D (DK)
R (REFUSED)

R1d. Why is the <MEASCAT1, MEASCAT2> equipment no longer installed [IF INSTALLED MEASURE IS OPERATIONAL; (IF EQUIP1, EQUIP2=1): or no longer operating] at this facility?

(RECORD VERBATIM RESPONSE)

(IF RESPONDENT WAS MOST INVOLVED IN THE DECISION AND MEASURE IS STILL OPERATING, ASK FREE RIDERSHIP QUESTIONS RELATED TO MEASCAT1, MEASCAT2)

(IF NOT PRIMARY DECISION MAKER FOR EITHER MEASURE, SKIP TO I1 AND DIAL THE MAIN DECISION MAKER IN R1b)

R3. Does your company have any corporate policies related to energy efficiency standards that you need to consider when purchasing new equipment or making improvements to this facility?

1 Yes
2 No [SKIP TO R6i]
D (DK) [SKIP TO R6i]
R (REFUSED) [SKIP TO R6i]

R4. Which of the following best describes this policy? (READ LIST)

1 Purchase energy efficient measures regardless of cost
2 Purchase energy efficient measures if it meets payback or return on investment criteria
3 Purchase standard efficiency measures that meet code
4 Something else (SPECIFY)
D (DK)
R (REFUSED)

R6i. (ASK IF PA DOES NOT HAVE TA INFORMATION) Did your company receive a technical assessment as part of your participation in the <PROGRAM>?

1 Yes [STUDY = Yes, STUDYTYPE = “technical assessment”]
2 No
D (DK)
R (REFUSED)

[IF NO <STUDY>, SKIP TO R9]

R6. <PA> paid <TA%> of the <TACOST> to conduct a <STUDY> at your facility to determine the cost-effectiveness of installing energy efficient <MEASCAT1 and MEASCAT2> project.

If <PA> had not paid a portion of the cost, would your company have paid <TACOST> to have a similar <STUDY> done at that same time?

1	Yes	[SKIP TO R9]
2	No	
D	(DK)	[SKIP TO R9]
R	(REFUSED)	[SKIP TO R9]

R7. Would you have paid to have the study done earlier than you did, at a later date, or never?

1 Earlier
2 Same time (REPEAT R6)
3 Later
4 Never
D (DK)
R (REFUSED)

R8. [IF R7 = EARLIER OR LATER (IF R7 = 1 OR 3)] How much [earlier/later] would you have had the study done?

___ YEARS (AND/OR) ___ MONTHS

D DK
R (REFUSED)

C2. [IF <PA> HAD NOT PAID A PORTION OF THE COST OF THE <STUDY>, COMPANY WOULD HAVE PAID FOR STUDY (R6=NO)] On a scale of 0 to 10, with 0 being no influence and 10 being a great deal of influence, how much influence did the information provided by the <STUDY> have on your decision to implement the [IF EFFICIENCY IS APPLICABLE; *IF EFF1, EFF2 = 1:* high efficiency] <MEASCAT1, MEASCAT2> project? (REPEAT FOR EACH MEASURE)

_____ (ENTER INFLUENCE RANKING)
D (DK)
R (REFUSED)

R9. Did you receive interest-free financing from <PA> which allowed you to pay for your portion of the project cost over time?

1 Yes
2 No
D DK

Free-Ridership

FR0. Please think back to the time when you were considering implementing the specific <MEASCAT1 and MEASCAT2> projects in <YEAR>.

What factors motivated your business to consider implementing new <MEASCAT1 and MEASCAT2> project(s)? (PROBE: What other factors did you consider?)

DO NOT READ LIST. PLEASE CHOOSE ALL THAT APPLY.

1 (Old equipment failed)
2 (Old equipment working poorly)
3 (Old equipment scheduled for replacement)
4 (Wanted to reduce maintenance costs)
5 (The incentive being offered through the program)
6 (The technical assistance offered through the program)
7 (Wanted to reduce energy bills)
8 (Wanted to save energy)
9 (Recommendation of third party contractor/engineer/design professional)
10 (Recommendation of <PA> staff)
11 (Recommendation of internal staff)
12 (Past experience with the program)
13 (Other - specify)
D (DK)
R (REFUSED)

START OF MEASURE LOOP

FR1-C9 will be asked of each measure category recalled that are still installed and operating - up to TWO measure categories.

INTRO3a

Now, I'd like to ask you about your decision to implement the <MEASCAT1> project. [IF THERE IS ALSO A SECOND MEASURE: Then, I'll repeat these questions for <MEASCAT2>].

INTRO3b

[IF SECOND MEASURE] Now I'd like to review the <MEASCAT2> project you implemented.

FR1. On a scale of 0 to 10, with 0 being not at all likely and 10 being very likely, how likely is it that your business would have implemented the same [IF QUANTITY IS GREATER THAN (IF QTY1, QTY2 = 1): quantity] [IF EFFICIENCY IS APPLICABLE (IF EFF1, EFF2 = 1): and efficiency of] <MEASCAT1, MEASCAT2> at that same time if the <PA> had not provided the <ALL ASSISTANCE>?

___ (0 TO 10)
D (DK)
R (REFUSED)

FR2. Did your company have any funds allocated to implement the <MEASCAT1, MEASCAT2> project before you talked with anyone about the program?

1	Yes	
2	No	[SKIP TO FR4]
D	(DK)	[SKIP TO FR4]
R	(REFUSED)	[SKIP TO FR4]

FR3a. Was it necessary to change the timing of the implementation, [IF QUANTITY IS GREATER THAN 1 (if QTY1, QTY2 = 1): the quantity of equipment] [IF EFFICIENCY IS APPLICABLE (IF EFF1, EFF2 = 1): or the efficiency level] of the <MEASCAT1, MEASCAT2> in order to qualify for the <PROGRAM>?

1	Yes	
2	No	[SKIP TO FR4]
D	(DK)	[SKIP TO FR4]
R	(REFUSED)	[SKIP TO FR4]

FR3b. What changes were necessary? [DO NOT READ; SELECT ALL THAT APPLY]

1 (Installation occurred SOONER than planned)
2 (Installation occurred LATER than planned)
3 (Installed MORE equipment than planned)
4 (Installed LESS equipment than planned)
5 (Equipment was MORE efficient than planned)
6 (Equipment was LESS efficient than planned)
7 (Removed MORE equipment than planned)
8 (Removed LESS equipment than planned)
9 (Other) (SPECIFY)
D (DK)
R (REFUSED)

FR4. Who was MOST responsible for actually recommending or specifying the [IF EFFICIENCY IS APPLICABLE (IF EFF1, EFF2 = 1): high efficiency] <MEASCAT1, MEASCAT2> project that was implemented through the <PROGRAM>?

DO NOT READ LIST, RECORD ONLY ONE

1 Respondent
2 Someone else in company (SPECIFY AND PROBE TO SEE IF SHOULD BE SPEAKING WITH THIS R)
3 Third-party design professional
4 Third-party engineer
5 Contractor
6 Manufacturer's representative
7 <PA> account manager
8 Someone else (SPECIFY)
D (DK)
R (REFUSED)

C1. [IF FR4= THIRD-PARTY DESIGN PROFESSIONAL, THIRD-PARTY ENGINEER, CONTRACTOR MANUFACTURER'S REPRESENTATIVE, OR <PA> ACCOUNT MANAGER (IF FR4=3, 4, 5, 6 OR 7)]

On a scale of 0 to 10, with 0 being no influence and 10 being a great deal of influence, how much influence did (FR4 response) have on your company's decision to implement the [IF EFFICIENCY IS APPLICABLE; *IF EFF1, EFF2 = 1:* high efficiency] <MEASCAT1,MEASCAT2> project so that it would qualify for the program?

_____ (ENTER INFLUENCE RANKING)
D (DK)
R (REFUSED)

FR5. I'd like to go over all the assistance you received from <PA>.

According to our records, the total cost for the project implemented at your facility in <DATE> through the <PROGRAM> was about <TOTCOST>. <PA> paid about <INC1, INC2> of the total cost of the [IF EFFICIENCY IS APPLICABLE; *IF EFF1, EFF2 = 1:* energy efficient] <MEASCAT1, MEASCAT2> project implemented through the program.

[IF NO <STUDY>: You may have also received some technical assistance from a <PA> rep, engineer, or equipment vendor.]

[IF <STUDY>: As I previously mentioned, <PA> paid <TACOST> for a <STUDY>.]

[IF <FINANCE> = Yes] <PA> also provided interest-free financing for up to 24 months for your portion of the project costs.

If <PA> had not paid a portion of the implementation cost OR provided any technical assistance or education [IF <FINANCE> = Yes: OR provided interest-free financing], would your business have implemented <u>any</u> <u>type</u> of <MEASCAT1, MEASCAT2> project <u>at the same time</u>?

1	Yes	[SKIP TO FR7a]
2	No	
D	(DK)	
R	(REFUSED)	

FR6a. Would you have implemented the <MEASCAT1, MEASCAT2> project earlier than you did, at a later date, or never?

1	Earlier	
2	Same time	[REPEAT FR5]
3	Later	
4	Never	[SKIP TO C3]
D	(DK)	[SKIP TO C3]
R	(REFUSED)	[SKIP TO C3]

FR6b. How much [earlier/later] would you have implemented the <MEASCAT1, MEASCAT2> project?

___	YEARS
___	MONTHS
D	DK
R	(REFUSED)

[IF QUANTITY IS NOT APPLICABLE FOR THIS MEASURE CATEGORY (IF QTY1, QTY2 = 0), SKIP TO FR8D]

FR7a. Without the program incentive, technical assistance, or financing, would your business have implemented the exact same quantity of <MEASCAT1, MEASCAT2> project [IF FR5=YES or DK: at that same time; IF FR5=2: within (TIMEFRAME IN FR6b)]?

1 Yes [SKIP TO FR8]
2 No
D (DK)
R (REFUSED)

[IF FR7A = DK OR R, SKIP TO C3]

FR7b. Compared to the amount of <MEASCAT1, MEASCAT2> that you implemented through the program, what percent of the project do you think your business would have purchased on its own during that timeframe?

(PROBE: Would you have purchased about one- fourth (25%), one-half (50%), three fourths (75%) of what you installed through the program?)

____ (ENTER PERCENTAGE: 1-99%)
D (DK)
R (REFUSED)

[IF EFFICIENCY IS NOT APPLICABLE FOR THIS MEASURE CATEGORY (IF EFF1, EFF2 = 0), SKIP TO C1]

FR8. You said your business would have installed [IF FR7A=YES: all; IF FR7A= NO: (FILL WITH FR7B %)] of the equipment on its own if the program had not been available. [ALL] Thinking about the <MEASCAT1, MEASCAT2> equipment you would have installed on your own, what percent of this equipment would have been . . . ?

(PROBE: Would about one-fourth (25%), one-half (50%), three fourths (75%) been of equal efficiency?)

a. of the same high efficiency as what was installed through the program?
____ (ENTER PERCENTAGE: 0-100%)
D (DK)

b. lower efficiency than what was purchased but higher than standard efficiency or code?
____ (ENTER PERCENTAGE: 0-100%)
D (DK)

c. standard efficiency or code
____ (ENTER PERCENTAGE: 0-100%)
D (DK)

(CHECK THAT THE THREE % SUM TO 100%; PROBE TO CLARIFY).

[IF QUANTITY IS GREATER THAN 1 (IF QTY1, QTY2 = 1), SKIP TO C1]

FR8d. Thinking about the <MEASCAT1, MEASCAT2> project you would have implemented on your own if the program had not been available, would it have been of the same high efficiency as what was installed through the program, lower efficiency than what was purchased but higher than standard efficiency, or standard efficiency or code?

1 Of the same high efficiency as what was installed through the program?
2 Lower efficiency than what was purchased but higher than standard efficiency
3 Standard efficiency or code
D (DK)
R (REFUSED)

C3. On a scale of 0 to 10, with 0 being no influence and 10 being a great deal of influence, how much influence did the <INC1,INC2> you received from <PA> have on your decision to implement the [IF EFFICIENCY IS APPLICABLE; *IF EFF1, EFF2 = 1:* high efficiency] <MEASCAT1,MEASCAT2> project?

_____ (ENTER INFLUENCE RANKING)
D (DK)
R (REFUSED)

Consistency Check Prompts

100% Free Ridership Consistency Check
[IF WOULD HAVE PURCHASED AT THE SAME TIME, IN THE SAME QUANTITY, AND OF THE SAM EFFICIENCY LEVEL; *IF FR5=1 AND FR7a=1 AND (FR8a=100% or FR8d = 1), ASK C4a-C7c*, ELSE SKIP TO C8]

C4a. Now I want to focus on what it would have cost your business to implement this project on its own without the program. On a scale of 0 to 10, with 0 being not at all likely and 10 being very likely, how likely is it that your business would have paid the additional <INC1,INC2> on top of the amount you already paid, to implement the same quantity and efficiency of <MEASCAT1,MEASCAT2> equipment at that same time?

___ (0 TO 10)
D (DK)
R (REFUSED)

C4b. (ASK IF C4a < 8) You said that you would have installed the same quantity and efficiency of equipment at that same time, but you also just said that there was a (FILL WITH C4a SCORE) in 10 likelihood of you paying the additional incentive provided by the <PA> program. Which of these is more accurate?

1 Installed same quantity & efficiency at same time [SKIP TO C9]
2 Likelihood of installing this without the program assistance was (C4a SCORE)
3 Something else (SPECIFY)

C5. How would your project have changed if <PROGRAM> had not contributed to the cost of the <MEASCAT1, MEASCAT2> project? (INDICATE ALL THAT APPLY) (DO NOT READ)

1 (Would not have changed) [SKIP TO C8]
2 (Would have postponed the project) (SPECIFY # MONTHS)
3 (Would have cancelled the project altogether)
4 (Would have repaired existing equipment)
5 (Kept using existing equipment)
6 (Purchased less efficient equipment) (ASK C7)
7 (Purchased fewer quantity) (ASK C6)
8 (Installed DIFFERENT type of equipment than planned) (SPECIFY)
9 (Other) (SPECIFY)
D (DK)
R (REFUSED)

C6. [IF C5=PURCHASED FEWER QUANTITY; *IF C5=7*) Compared to the amount of <MEASCAT1, MEASCAT2> that you implemented through the program, what percent do you think your business would have purchased on its own at that same time? (PROBE: Would you have purchased about one- fourth (25%), one-half (50%), three fourths (75%) of what you installed through the program?)

____ (ENTER PERCENTAGE: 1-99%)
D (DK)
R (REFUSED)

C7. [IF C5=PURCHASED LESS EFFICIENT EQUIPMENT; *IF C5=6*) Thinking about the equipment you would have implemented on your own, what percent of this equipment would have been . . . ? (PROBE: Would about one-fourth (25%), one-half (50%), three fourths (75%) been of equal efficiency?)

a. of the same high efficiency as what was installed through the program?
____ (ENTER PERCENTAGE: 0-100%)
D (DK)

b. lower efficiency than what was purchased but higher than standard efficiency or code?
____ (ENTER PERCENTAGE: 0-100%)
D (DK)

c. standard efficiency or code
____ (ENTER PERCENTAGE: 0-100%)
D (DK)

(CHECK THAT THE THREE % SUM TO 100%; PROBE TO CLARIFY).

0% Free Ridership Consistency Check

C8 (IF SMALL BUSINESS - ASK IF AT LEAST SOMEWHAT LIKELY TO HAVE INSTALLED THE MEASURE WITHOUT THE PROGRAM BUT LATER STATES WOULD HAVE WAITED AT LEAST TWO YEARS (FR1 > 3 AND FR6b > **24** MONTHS OR NEVER)

(IF MED/LARGE C&I - ASK IF AT LEAST SOMEWHAT LIKELY TO HAVE INSTALLED THE MEASURE WITHOUT THE PROGRAM BUT LATER STATES WOULD HAVE WAITED AT LEAST FOUR YEARS (FR1 > 3 AND FR6b > **48** MONTHS OR NEVER)

Earlier in the interview, you said there was a (FR1 SCORE) in 10 likelihood that you would have implemented the same quantity and efficiency of <MEASCAT1, MEASCAT2>equipment at that same time in the absence of the program assistance. But you also said you would not have implemented the <MEASCAT1, MEASCAT2> project within <2/4> years of when you did. Which of these is more accurate?

1 The likelihood of installing this without the program assistance was (FR1 SCORE)
2 Would not have installed anything within 2/4 years
3 Something else (SPECIFY)
D (DK)
R (REFUSED)

Additional Consistency Check

C9. (IF 100% FREE-RIDER; *IF FR5=1 AND FR7a=1 AND (FR8a=100% or FR8d = 1) AND C4b = 1 AND (C2 > 6 OR C3 > 6*)) **PROMPT**: "Previously you stated that you would have installed the exact same equipment at the same time without the program. But, you also stated that the …

(IF C2 > 6 FILL: program-sponsored study)
(IF C3 > 6 FILL: program incentive and financing options)
(IF C2 > 6 & C3 > 6 FILL: program-sponsored study, incentive, and financing options)

… was influential in your decision.)

(IF 0% FREE-RIDER: *IF FR6a = NEVER OR DK AND (C2 < 5 OR C3 < 5)* **PROMPT**: "Previously you stated that you would not have installed any equipment without the program. You also stated that the …

(IF C2 < 5 FILL: program-sponsored study)
(IF C3 < 5 FILL: program incentive and financing options)
(IF C2 < 5 & C3 < 5 FILL: program-sponsored study, incentive, and financing options)

… was not influential in your decision.)

(ASK OF ALL) I'd like to better understand your purchase decision. In your own words, please describe what impact, if any, all the assistance you received through the program had on your decision to install the amount of energy efficient <MEASCAT1, MEASCAT2> equipment at the time you did?

(RECORD VERBATIM RESPONSE)

SKIP1

(REPEATS QUESTIONS BEGINNING FROM INTRO3B FOR SECOND MEASURE – IF NO OTHER MEASURES – CONTINUE)

[IF MEAS2 = 1 GO TO INTRO3B]
[IF MEAS2 = 0 GO TO PP1]

Impact of Previous Program Participation

[IF NEVER WOULD HAVE INSTALLED OR ALL EQUIPMENT WOULD HAVE BEEN OF STANDARD EFFICIENCY AND UNLIKELY TO HAVE PURCHASED WITHOUT PROGRAm ((IF FR6A = NEVER OR FR8A = 0% OR FR8D <> 1) AND FR1 < 4) SKIP TO COM]

PP1. Had your business participated in <PA>'s <PROGRAM> before you implemented the energy efficient project in <DATE>?

1 Yes
2 No [SKIP TO S1a]
D (DK) [SKIP TO S1a]
R (REFUSED) [SKIP TO S1a]

PP2. On a scale of 0 to 10, with 0 being 'not at all important and 10 being 'very important', how important was your previous experience with a <PA> program when making the decision to implement the <MEASCAT1, MEASCAT2> project at this facility around <DATE>?

0 – 10
D (DK)

PP3. I'm going to read you several statements. For each statement, please tell me whether you agree or disagree that this statement applies to your business. There are no right or wrong answers; we just want your honest opinion.
(REPEAT IF NECESSARY)

1 Agree
2 Disagree
D (DK)
R (REFUSED)

Our previous experience implementing energy efficient projects through the <PROGRAM> . . . .

a. Has made our firm more likely to consider energy efficient equipment
b. Has made our firm more likely to install energy efficient equipment
c. Has given us more confidence in the financial benefits of energy efficient equipment
d. Has given us more confidence in the nonfinancial benefits of energy efficient equipment

Like Spillover[37]

START OF MEASURE LOOP
S1a-S4b will be asked of each measure category recalled - up to TWO measure categories.

S1a. Now I'd like you to think of the time since you participated in the <PROGRAM> in <DATE>.

Has your company implemented any <MEASCAT1, MEASCAT2> projects for this or other facilities in Massachusetts **on your own,** that is without a rebate from <PA>?

1 Yes
2 No [SKIP TO SKIP2]
D (DK) [SKIP TO SKIP2]

[IF EFFICIENCY IS NOT APPLICABLE; *IF EFF1, EFF2 = 0, SKIP TO S2a*]

S1b. Was this equipment of **the same efficiency level or a higher level of efficiency** as the equipment you installed through the program?

1 Yes [SKIP TO S2a]
2 No
D (DK)

S1c. Was this equipment more energy efficient than standard efficiency or code equipment?

1 Yes
2 No [SKIP TO SKIP2]
D (DK) [SKIP TO SKIP2]

S2a. Thinking of the <MEASCAT 1, MEASCAT 2> equipment that you installed on your own, how does the quantity compare to what you installed through the program? Did you install more, less or the same amount of <MEASCAT 1, MEASCAT 2>?

(PROBE: We're looking for a percent compared to the amount installed through the program. For example, was it about one- fourth of what you installed through the

[37] As these surveys are being conducted soon after implementation, estimates of like and unlike spillover are likely to be limited as participants have not had adequate time to install additional equipment. Participant spillover will be asked at a later date as part of the C&I NEI study.

program, one-half of what you installed through the program, the same (100%) amount as you installed through the program, twice as much as what you installed through the program (200%) or some other amount?)

1 More (How much more? Enter percentage: 0-100%)
2 Less (How much less? Enter percentage: 0-100%)
3 Same
D (DK)

S2b. [IF S2a <> SAME AMOUNT OF <MEASCAT 1, MEASCAT 2>; *IF S2a <> 3*]So the additional energy efficient equipment you bought on your own was <percentage from S2a> as much as you got through the program?

1 Yes
2 No [correct S2a]

S3a. Did a recommendation by the contractor, engineer, or designer who you worked with under the <PROGRAM> influence your decision to implement some or all of this [IF EFFICIENCY IS APPLICABLE; (IF EFF1, EFF2 = 1): efficient] <MEASCAT1, MEASCAT2> equipment on your own?

1 Yes
2 No
D (DK)
R (REFUSED)

S3b. Did your experience with the energy efficient projects implemented through the <PROGRAM> influence your decision to implement some or all of this [IF EFFICIENCY IS APPLICABLE; (IF EFF1, EFF2 = 1): efficient] <MEASCAT1, MEASCAT2> equipment on your own?

1 Yes
2 No
D (DK)
R (REFUSED)

S3c. Did your participation in any past program offered by <PA> influence your decision to implement some or all of this [IF EFFICIENCY IS APPLICABLE; (IF EFF1, EFF2 = 1): efficient] <MEASCAT1,MEASCAT2> equipment on your own?

1 Yes
2 No
D (DK)
R (REFUSED)

S3d. On a scale of 0 to 10, where 0 is “no influence at all” and 10 is “a great deal of influence”, how much influence did your participation in the <PA> program have on your decision to install this equipment without an incentive?

__ 0-10 rating
D (DK)

S4a. Why didn't you implement this <MEASCAT1, MEASCAT2> project through a <PA> program?

[DO NOT READ - SELECT ALL THAT APPLY]

1 (Too much paperwork)
2 (Cost savings not worth the effort of applying)
3 (Takes too long for approval)
4 (The equipment would not qualify)
5 (Vendor does not participate in program)
6 (Outside <PA>’s service territory)
7 (No time - needed equipment immediately)
8 (Thought the program ended)
9 (Didn't know the equipment qualified under another program)
10 (Just didn't think of it)
11 (Unable to get rebate--unsure why)
12 (Other) (SPECIFY)
D (DK)

S4b. [IF S4a = THE EQUIPMENT WOULD NOT QUALIFY; *IF S4a = 4*) Why wouldn't the equipment qualify?

(RECORD VERBATIM RESPONSE)

SKIP2

(REPEATS SPILLOVER QUESTIONS FOR SECOND MEASURE – IF NO OTHER MEASURES – CONTINUE)

[IF MEAS2 = 1 GO TO S1A]
[IF MEAS2 = 0 GO TO S5]

Unlike Spillover[38]

S5. Since participating in <PROGRAM>, had your company purchased, installed, or implemented any other type of energy efficiency equipment **on your own,** that is without a rebate from <PA>?

1 Yes
2 No [SKIP TO NE1]
D (DK) [SKIP TO NE1]

S6. What did you install?

Record type: _____
Record quantity: _____
Record size or capacity: _____

S7a. Would this project have qualified for an incentive through the <PROGRAM>?

1 Yes
2 Yes, implemented through a program [SKIP TO NE1]
2 No [SKIP TO NE1]
D (DK) [SKIP TO NE1]

S7b. Did a recommendation by the contractor, engineer, or designer who you worked with under the <PROGRAM> influence your decision to implement some or this equipment on your own?

1 Yes
2 No
D (DK)
R (REFUSED)

S7c. Did your experience with the energy efficient projects implemented through the <PROGRAM> influence your decision to implement some or this equipment on your own?

1 Yes
2 No
D (DK)
R (REFUSED)

[38] More detailed unlike spillover questions will be developed as part of the C&I NEI study.

S7d. Did your participation in any past program offered by <PA> influence your decision to implement some or all of this equipment on your own?

1 Yes
2 No
D (DK)
R (REFUSED)

S7e. On a scale of 0 to 10, where 0 is “no influence at all” and 10 is “a great deal of influence”, how much influence did your participation in the <PA> program have on your decision to install this equipment without an incentive?

__ 0-10 rating
D (DK)

S8a. Why didn't you implement this project through a <PA> program?

DO NOT READ - SELECT ALL THAT APPLY

1 (Too much paperwork)
2 (Cost savings not worth the effort of applying)
3 (Takes too long for approval)
4 (The equipment would not qualify)
5 (Vendor does not participate in program)
6 (Outside <PA>’s service territory)
7 (No time - needed equipment immediately)
8 (Thought the program ended)
9 (Didn't know the equipment qualified under another program)
10 (Just didn't think of it)
11 (Unable to get rebate--unsure why)
12 (Other) (SPECIFY)
D (DK)

S8b. [IF S8a = EQUIPMENT WOULD NOT QUALIFY (IF S8a = 4)] Why wouldn't the project qualify?

(RECORD VERBATIM RESPONSE)

Expected NEI

NE1. Prior to participating in the program, did you expect any impacts other than energy savings?

1 Yes
2 No [SKIP TO COM]
D (DK) [SKIP TO COM]

NE2. Did you view these effects as a negative or positive benefit?

1 Negative [SKIP TO COM]
2 Positive
D (DK)

NE3. What were the positive benefits? (SELECT ALL THAT APPLY)

1 Sales
2 Production/productivity
3 Equipment life
4 Maintenance costs
5 Waste generation
6 Personnel needs
7 Injury or illness
8 Other (SPECIFY)

NE4. [IF POSITIVE BENEFIT, *NE2 = 2*] Did the expected positive benefits influence your decision to participate in the program?

1 Yes
2 No
D (DK)

NE5. Did the program influence your expectations of the positive benefits?

1 Yes
2 No
D (DK)

Wrap-up

COM. Do you have any comments or suggestions for the program?

(RECORD VERBATIM RESPONSE)

QRNAME.
For verification purposes, would you spell your first and last name for me?

(RECORD VERBATIM RESPONSE)

CLARIFY.

If we would need to clarify some of the information I asked you, would it be alright if we called you back?

1 Yes
2 No

A4. [ASK IF C1 > 6]
We would like to talk to the person who was most influential in recommending or specifying the efficient <MEASCAT1, MEASCAT2> equipment to install through the program. Earlier you mentioned that this was [FILL WITH FR4 RESPONSE]. Could you give me the name and telephone number of this person?

1 Yes (Record contact information)
2 No, REFUSED to give this information
3 No, no outside advisor involved
4 [IF SECOND MEASURE] (SAME CONTACT INFO AS PREVIOUS MEASURE)
D (DK)

END

Those are all the questions I have for you. I'd like to thank you for your time with this important evaluation.

2010 Influential Design Professional/Vendor Free-Ridership Survey

Variable List

<CONTACT>	Customer Contact Name
<CUST>	Customer/Facility Name
<ADDR>	Service address where equipment was installed
<MEASCAT1, MEASCAT2>	End-use Category (i.e. lighting)
<MEASCAT1a-MEASCAT1h>	Detailed measure descriptions
<MEASCAT2a-MEASCAT2h>	Detailed measure descriptions
<TA>	"1" if a Technical Assessment Study was conducted
<TA%>	Percent of TA study paid by utility/sponsor (by program)
<TACOST>	Total cost paid by utility/sponsor for TA study (by program)
<INC1, INC2>	Utility/sponsor incentive for Measure categories
<QTY1, QTY2>	0=quantity is not applicable for this measure category (measure qty = 1 or quantity is not relevant as in delamping, recycling), 1=quantity greater than 1
<EFF1, EFF2>	0=efficiency is not applicable for this measure category (e.g., insulation, VFD, delamping, recycling), 1=efficiency is applicable
<EQUIP1, EQUIP2>	0 if installed measure isn't equipment that is operational (e.g., insulation), 1=if installed measure is operational
<TOTCOST>	Total project cost (customer cost+utility cost) for an account (by program)
<PROGRAM>	Utility/sponsor programs the vendor has been involved with
<PA>	Utility/sponsor name

Procedure

The customer-identified vendors will be exported from each PA study and combined into a single sample file. This file will be checked for missing contact information and we will fill in phone numbers where possible. Cases will then be sorted by company, contact, and phone number to identify "multiples". Cases with the same contact names will be called together and the contact will be alerted that they have been referred by more than one customer. This set of sample cases will receive the free-rider questions only.

Introduction

INTRO
Hello, my name is __, and I am calling on behalf of <**PA** >. We are talking with some of the design professionals and contactors who were involved with the <**PROGRAM**> in 2010. I'm not selling anything; I'd just like to ask you about the types of equipment that your firm recommended, sold, or installed through this/these program(s) in 2010.

Before we start, I would like to inform you that for quality control purposes, this call will be recorded and monitored.

(Timing: This survey will take less than 15 minutes of your time. IF NOT A GOOD TIME, SET UP CALL BACK APPOINTMENT OR OFFER TO LET THEM CALL US BACK AT 1-800-454-5070)

(Sales concern: I am not selling anything. Your responses will be kept confidential by our firm and the <**PA**>. If you would like to talk with someone from there, you can call [CONTACT NAME AND PHONE NUMBER FOR SPONSORS INCLUDED IN THIS CALL].

Free-Ridership Questions

INTRO2
I'd like to review the <**MEASCAT1**, **MEASCAT2**> you recommended or specified through the <**PROGRAM**> for <**PA**>.

VR1 Do you recall recommending <**MEASCAT1**>, which included <**DESC1**> for <**CUST**> at <**ADDR**> through the <**PROGRAM**> in 2010?

1 Yes [SKIP TO V1a]
2 No
3 This equipment was never installed [IF NUMBER OF MEASURE CATEGORIES=2, SKIP TO VR2; ELSE SKIP TO END]
D (DK)
R (Refused)

VR1a Is there someone else at your firm who would be more familiar with this equipment?

1 Yes - Continue [ENTER CONTACT INFO & TRANSFER. GO THROUGH INTERVIEW WITH OTHER CONTACT IF AVAILABLE, OTHERWISE SET CALLBACK AND UPDATE CONTACT INFORMATION.]
2 Yes – Not available [ENTER CONTACT INFO & EXIT]
3 No – Continue
4 Contact no longer with the company

V1a First I'd like to ask you about your decisions to recommend <**MEASCAT1**> through the <**PROGRAM**>. Were you involved in the decision-making process at the design stage when the <**MEASCAT1**> equipment was specified and agreed upon for this facility?

1 Yes [IF # OF MEASURE CATEGORIES = 2, SKIP TO VR2, ELSE SKIP TO INTRO3a]
2 No
D (DK)

V1b At what point in the process did you become involved?

(RECORD VERBATIM RESPONSE)
(DK)
(REFUSED)

V1c What was your role?

(RECORD VERBATIM RESPONSE)
(DK)
(REFUSED)

[IF NO SECOND MEASURE, SKIP TO INTRO3a]

VR2 Do you recall recommending <**MEASCAT2**> which included <**DESC2**> for <**CUST**> at <**ADDR**> through the <**PROGRAM**> in 2010?

1 Yes [SKIP TO V2a]
2 No
3 This equipment was never installed [SKIP TO INTRO3a IF INSTALLED MEASURE CATEGORY 1; ELSE SKIP TO END]
D (DK)

VR2a Is there someone else at your firm who would be more familiar with this equipment?

1 Yes - Continue [ENTER CONTACT INFO & TRANSFER IF NOT CONTACT FOR MEASURE 1]
2 Yes – Not available [ENTER CONTACT INFO & EXIT IF NOT CONTACT FOR MEASURE 1]
3 No – Continue
4 Contact no longer with the company

[IF DIDN'T RECALL MEASURE 1, MEASURE 1 WAS NOT INSTALLED, OR R WAS NOT THE CONTACT FOR MEASURE 1, SKIP TO END; ELSE SKIP TO INTRO3a AND ONLY ASK QUESTIONS FOR MEASURE 1]

V2a Were you involved in the decision-making process at the design stage when the <**MEASCAT2**> equipment was specified and agreed upon for this facility?

1 Yes
2 No
D (DK)

V2b At what point in the process did you become involved?

(RECORD VERBATIM RESPONSE)
(DK)
(REFUSED)

V2c What was your role?

(RECORD VERBATIM RESPONSE)
(DK)
(REFUSED)

[IF TA=0 SKIP TO INTRO3a]

VP1a According to our records, <**PA**> paid <**TA**%> of the <**TACOST**> to conduct a <**STUDY**> for <**CUST**> to determine the cost-effectiveness of installing <**MEASCAT1**, **MEASCAT2**> equipment.

If <**PA**> had not paid a portion of the cost, do you think <**CUST**> would have paid <**TACOST**> to have a similar <**STUDY**> done at the same time?

1 Yes
2 No
D (DK)

VR9 To the best of your knowledge, did <CUSTOMER> receive interest-free financing from <PA> which allowed them to pay for their portion of the project cost over time?

1 Yes
2 No

[INTERVIEWER: START OF MEASURE LOOPS. VA3 THROUGH VF10 WILL BE ASKED OF EACH MEASURE CATEGORY RECALLED - UP TO TWO MEASURES.]

INTRO3a [FIRST MEASURE]
Now, I'd like to ask you some questions about your decision to recommend <**MEASCAT1**> equipment. [IF THERE IS ALSO A SECOND MEASURE: Then, I'll repeat these questions for <**MEASCAT2**> equipment.]

INTRO3b [IF SECOND MEASURE]
Now I'd like to review the <**MEASCAT2**> equipment you recommended.

VA1 On a scale of 0 to 10, with 0 being no influence and 10 being a great deal of influence, how much influence did your firm have on specifying the efficiency levels or features of <**MEASCAT1, MEASCAT2**> so that it would qualify for the program?

__ (0-10)
D (DK)

(IF VA1 < 7 AND NO OTHER MEASURE, SKIP TO END; IF VA3<7 AND ANOTHER MEASURE CATEGORY, REASK VA1 OF SECOND MEASURE CATEGORY; ELSE SKIP TO VP1a)

VPI The next set of questions ask about <**CUST**>'s planning and installation decisions through <**PROGRAM**> in 2010.

VP1a As far as you know, did <**CUST**> have funds allocated to install any of this equipment before you talked with them about the program?

1 Yes
2 Yes, but don't remember specifics
3 No [SKIP TO ATXT3]
D (DK) [SKIP TO ATXT3]
R (Refused) [SKIP TO ATXT3]

VP1b (IF YES) What plans existed?

(RECORD VERBATIM RESPONSE)
(DK)
(REFUSED)

VP2a Was it necessary to change the timing of the installation, the quantity of equipment installed or the efficiency level of the <**MEASCAT1**, **MEASCAT2**> equipment installed in order to qualify for the <**PROGRAM**>?

1 Yes
2 Yes, but don't remember specifics [SKIP TO ATXT3]
3 No [SKIP TO ATXT3]
D (DK) [SKIP TO ATXT3]
R (Refused) [SKIP TO ATXT3]

VP2b What changes were necessary? [INDICATE ALL THAT APPLY]

1 (Installation occurred SOONER than planned)
2 (Installation occurred LATER than planned)
3 (Installed MORE equipment than planned)
4 (Installed LESS equipment than planned)
5 (Equipment was MORE efficient than planned)
6 (Equipment was LESS efficient than planned)
7 (Other - specify)
D (Don't know)
R (Refused)

VATXT3
According to our records, the total cost for all equipment installed at <**CUST**>'s facility was about <**TOTCOST**>. <**PA**> paid about <**INC1**, **INC2**> of the total cost of the <**MEASCAT1**, **MEASCAT2**>.

<**CUST**> may have also received some technical assistance from <**PA**> or a contribution toward the cost of a technical assessment study.

VF1 If <**PA**> had not paid a portion of the implementation cost, would your company have recommended or specified any type of <**MEASCAT1**, **MEASCAT2**> equipment to <**CUST**> at the same time?

1 Yes
2 No [SKIP TO VC2]
D (DK) [SKIP TO VC2]

[IF QTY1, QTY2 = 0, SKIP TO VF3d]

VF2a Without the program incentive, technical assistance, or education, would your company have recommended or specified the exact same quantity of <**MEASCAT1**, **MEASCAT2**> for <**CUST**> at the same time?

1 Yes [SKIP TO VF3]
2 No
D (DK)

VF2b Compared to the amount that you recommended through the program, what percentage of the overall quantity of <**MEASCAT1**, **MEASCAT2**> equipment do you think your company would have recommended or specified without assistance from <**PA**>?

(PROBE: Would you have recommended/specified about one-fourth (25%), one-half (50%), three fourths (75%) of what was installed through the program?)

____ ENTER PERCENTAGE (0-100%, 998=DK)

[IFVF2b =0, SKIP to VC2]
[IF EFF1, EFF2 = 0, SKIP TO VC2]

VF3 You said you would have recommended or specified [IF VF2a=1: all the] [IF VF2a=2 OR D SHOW: at least some] <**MEASCAT1**, **MEASCAT2**> for <**CUST**> if the program had not been available.

What percent of the equipment that you would have recommended would have been…

a. of the same high efficiency as what was installed through the program?

____ (ENTER PERCENTAGE: 0-100%)
D (DK)

b. lower efficiency than what was purchased but higher than standard efficiency or code?

____ (ENTER PERCENTAGE: 0-100%)
D (DK)

c. standard efficiency or code?

____ (ENTER PERCENTAGE: 0-100%)
D (DK)

[IF QTY1, QTY2 = 1, SKIP TO VC2]
[IF EFF1, EFF2 = 0, SKIP TO VC2]

VF3d Thinking about the <**MEASCAT1**, **MEASCAT2**> equipment you would have recommended if the program had not been available, would it have been of the same high efficiency as what was installed through the program, lower efficiency than what was purchased but higher than standard efficiency, or standard efficiency or code?

1 Of the same high efficiency as what was installed through the program?
2 Lower efficiency than what was purchased but higher than standard efficiency
3 Standard efficiency or code
D (DK)
R (REFUSED)

VC2 [IF STUDY] On a scale of 0 to 10, with 0 being no influence and 10 being a great deal of influence, how much influence did the information provided by the <**STUDY**> have on your decision to recommend the [IF **EFF1**, **EFF2** = 1: high efficiency] <**MEASCAT1**,**MEASCAT2**> project?

_____ (ENTER INFLUENCE RANKING)
D (DK)
R (REFUSED)

VC3 On a scale of 0 to 10, with 0 being no influence and 10 being a great deal of influence, how much influence did the <**INC1**,**INC2**> <**CUST**> received from <**PA**> have on your decision to recommend the [IF **EFF1**, **EFF2** = 1:high efficiency] <**MEASCAT1**,**MEASCAT2**> project?

_____ (ENTER INFLUENCE RANKING)
D (DK)
R (REFUSED)

(IF VF1=1 AND VF2a=1 AND VF3a=100%, ASK VF4-VF7; ELSE SKIP TO VF8)

VF4a Now I want to focus on what it would have cost <**CUST**> to install this equipment on its own without the program. On a scale of 0 to 10, with 0 being not at all likely and 10 being very likely, how likely would they have been to pay the additional <**INC1**,**INC2**> on top of the amount they already paid, to implement the same quantity and efficiency of <**MEASCAT1**, **MEASCAT2**> equipment at that same time?

___ (0 TO 10)
D (DK)
R (REFUSED)

(IF VF4 > 7 SKIP TO VF8)

VF5 How would their project have changed if <**PROGRAM**> had not contributed to the cost of the <**MEASCAT1**, **MEASCAT2**>?
(INDICATE ALL THAT APPLY) (DO NOT READ)

1 Would not have changed [SKIP TO VF8]
2 (Would have postponed the project) (SPECIFY # MONTHS)
3 (Would have cancelled the project altogether)
4 (Would have repaired existing equipment)
5 (Kept using existing equipment)
6 (Purchased less efficient equipment) (ASK VF7)
7 (Purchased fewer quantity) (ASK VF6)
8 (Installed DIFFERENT type of equipment than planned) (SPECIFY)
9 (Other) (SPECIFY)
D (DK)
R (REFUSED)

VF6 (IF VF5=7) Compared to the amount of <**MEASCAT1**, **MEASCAT2**> that <**CUST**> implemented through the program, what percent do you think they would have purchased on their own at that same time?

(PROBE: Would you have purchased about one- fourth (25%), one-half (50%), three fourths (75%) of what you installed through the program?)

____ (ENTER PERCENTAGE: 0-99%)
D (DK)
R (REFUSED)

(IF VF6 = 0 SKIP TO VF8)
(IF QTY1, QTY2 = 0 SKIP TO VF8)

VF7 (IF VF5=6) Thinking about the equipment <**CUST**> would have implemented on your own, what percent of this equipment would have been . . . ?

(PROBE: Would about one-fourth (25%), one-half (50%), three fourths (75%) been of equal efficiency?)

a. of the same high efficiency as what was installed through the program?

____ (ENTER PERCENTAGE: 0-100%)
D (DK)

b. lower efficiency than what was purchased but higher than standard efficiency or code?
____ (ENTER PERCENTAGE: 0-100%)
D (DK)

c. standard efficiency or code
____ (ENTER PERCENTAGE: 0-100%)
D (DK)

(CHECK THAT THE THREE % SUM TO 100%; PROBE TO CLARIFY).

VF8 On a scale of 1 to 10, with 1 being 'not at all important and 10 being 'very important', how important was your previous experience with a <**PA**> program when making the decision to recommend or install <**MEASCAT1**, **MEASCAT2**>for this customer?

_____
D (DK)
N NA – No previous program experience

VF9 (IF VF1=1 AND VF2a=1 AND (VF3a=100% or VF3d = 1) AND VF5 = 1 AND (VC2 > 6 OR VC3 > 6) PROMPT: "Previously you stated that <**CUST**> would have recommended the exact same equipment at the same time without the program. But, you also stated that the …

(IF VC2 > 6 FILL: program-sponsored study)
(IF VC3 > 6 FILL: program incentive)
(IF VC2 > 6 & VC3 > 6 FILL: program-sponsored study and incentive)

… was influential in your decision to make the recommendations that you did.)

(IF VF1 = NO OR DK AND (VC2 < 5 OR VC3 < 5) PROMPT: "Previously you stated that <**CUST**> would not have installed any equipment without the program. You also stated that the …

(IF VC2 < 5 FILL: program-sponsored study)
(IF VC3 < 5 FILL: program incentive)
(IF VC2 < 5 & VC3 < 5 FILL: program-sponsored study and incentive)

… was not influential in their decision.)

I'd like to better understand <**CUST**>'s purchase decision. Please describe what impact, if any, the program had <**CUST**>'s decision to install the energy efficient <**MEASCAT1**,**MEASCAT2**> equipment at the time they did?

(RECORD VERBATIM RESPONSE)
(DK)
(REFUSED)

END We are almost finished calling customers about their experience with the program. If another customer identifies you as being influential in their decision to install energy efficient equipment, would it be alright for us to call you back for just a couple of questions?

1 YES
2 NO

VRNAME

For verification purposes, would you spell your first and last name for me?

(RECORD VERBATIM RESPONSE)

COMMENTS

That is all the questions I have for you. Thank you for your participation. Do you have any comments?

(RECORD VERBATIM RESPONSE)

2010 Design Professional/Vendor Nonparticipant Like Spillover Survey

Variable List

<CONTACT>	Customer Contact Name
<CUST>	Customer/Facility Name
<ADDR>	Service address where equipment was installed
<PA>	Sponsors the vendor has worked with on energy efficiency projects
<PROGRAM>	Utility/sponsor programs the vendor has been involved with
<ME1-ME14>	Types of equipment specified/sold as part of spillover questions

Procedure

The vendors identified in the sponsor databases will be asked the nonparticipant spillover questions. We will focus on reaching the contacts listed in the database.

Introduction

INTRO

Hello, my name is __, and I am calling on behalf of <**PA**>. We are talking with some of the design professionals and contactors who were involved with the <**PROGRAM**> in 2010. I'm not selling anything; I'd just like to ask you about the types of equipment that your firm recommended, sold, or installed through this/these program(s) in 2010.

Before we start, I would like to inform you that for quality control purposes, this call will be recorded and monitored.

(Timing: This survey will take less than 15 minutes of your time. IF NOT A GOOD TIME, SET UP CALL BACK APPOINTMENT OR OFFER TO LET THEM CALL US BACK AT 1-800-454-5070)

(Sales concern: I am not selling anything. Your responses will be kept confidential by our firm and <**PA**>. If you would like to talk with someone from there, you can call [CONTACT NAME AND PHONE NUMBER FOR SPONSORS INCLUDED IN THIS CALL].

VNP1a Our records show that your firm specified, sold, and/or installed <**MEx**> to commercial and industrial customers in 2010 through the <**PROGRAM**>. This includes equipment such as <**DESC**>.
Is that correct?
[INTERVIEWER: PLEASE VERIFY EACH TYPE OF EQUIPMENT THAT SHOWS FOR THE VENDOR]

1 Yes
2 No
D (DK)
ME2 = Motors
ME2a = Motors: New
ME2b = Motors: Failed/Stock
ME3 = HVAC equipment
ME3a = HVAC: Unitary
ME3b = HVAC: Other
ME4 = Variable speed drives
ME5 = Lighting equipment
ME6 = Non-Lighting equipment
ME7 = Transformers
ME8 = Compressed air
ME9 = Refrigeration
ME10 = Process equipment and system
ME11 = Process cooling equipment
ME12 = VSDs on non-HVAC systems
ME13 = Comprehensive Chillers
ME14 = Equipment converting electric DHW to gas, Comprehensive design projects, O&M projects

Note: The measure categories listed above will closely match measure categories as defined in the customer sample. When asking vendors about each measure category, we will reference the specific measure-level descriptions noted in the database.

[VNP1b-NP8 WILL BE ASKED FOR EACH MEASURE WHERE MEx=1 where x=measure category number defined above].

VNP1b Prior to selling equipment that received a rebate from the <**PA**> program, in what percentage of your commercial projects did you install high efficiency <**MEx**>?

__ ENTER PERCENTAGE
D (DON'T KNOW)
R (REFUSED)

VNP1c And during the past year, in what percentage of your commercial projects did you install high efficiency <**MEx**>?

__ ENTER PERCENTAGE
D (DON'T KNOW)
R (REFUSED)

VNP2 Please think about all the program-eligible <**MEx**> you specified, sold and/or installed for <**PA**> customers in 2010.
Did you specify, sell and/or install any of this program-eligible <**MEx**> to customers of <**PA**> <u>without</u> the customer participating in a <**PA**> program??

1 Yes
2 No [SKIP TO NEXT CATEGORY]
D (DK) [SKIP TO NEXT CATEGORY]

VNP3 (IF VNP2 = Yes) What percent of all of this program-eligible <**MEx**> you specified, sold and/or installed for <**PA**> customers in 2010 did <u>not</u> receive an incentive through a <**PA**> program?

_____%
D (DK)

(ASK VNP4-VNP7 OF EACH MEASURE WHERE VNP3 > 0%)

VNP4 In 2010, you mentioned that about [___%] of the <**MEx**> you specified and/or installed would have been eligible for an incentive through a <**PA**> program, but did not receive an incentive.
What are the main reasons why your firm did not request a customer incentive for this energy saving equipment you specified/installed?
(DO NOT READ—INDICATE ALL THAT APPLY; PROBE, WHAT ELSE?)

1 not worth the paperwork for our firm to help the customer apply for the incentive
2 customer did not want the hassle of applying for the incentive
3 takes too long for approval
4 reached the maximum amount I could install through the program
5 the equipment would not qualify→Why not? _____
6 vendor does not participate in program
7 outside [retail company] service territory
8 no time – needed equipment immediately
9 thought the program ended
10 didn't know the equipment qualified under another program
11 just didn't think of it
12 unable to get rebate (unsure why)
13 other *(SPECIFY)*
14 (DK)

VNP5 I'm going to read you 3 statements. For each statement, please tell me whether you agree or disagree that this statement applies to your company. There are no right or wrong answers; we just want your honest opinion.
Our past experience specifying or installing <**MEx**> through energy efficiency programs has convinced us that this equipment is cost effective or beneficial even without a program incentive.

0 Disagree
1 Agree

VNP6 We are better able to identify opportunities to improve energy efficiency by using high efficiency <**MEx**> because of our previous experience with the performance of energy efficient equipment installed through energy efficiency programs, and what we learned through working with <PA>.

0 Disagree
1 Agree

VNP7 We are more likely to discuss energy efficient options with all of our customers when developing project plans for <**MEx**> because of our previous experience with the performance of energy efficient equipment installed through energy efficiency programs, and what we learned through working with <**PA**>.

0 Disagree
1 Agree

VNPX Please describe what impact, if any, the <**PROGRAM**> had on your decision to specify or install energy efficient <**MEx**> outside of the program.

(RECORD VERBATIM RESPONSE)

END We are almost finished calling customers about their experience with the program. If a customer identifies you as being influential in their decision to install energy efficient equipment, would it be alright for us to call you back for just a couple of questions?

1 YES
2 NO

VRNAME
For verification purposes, would you spell your first and last name for me?

COMMENTS
That is all the questions I have for you. Thank you for your participation. Do you have any comments?

APPENDIX F: SCORING FLOWCHARTS

Figure F-1. 2010 Free-ridership Scoring Flowchart

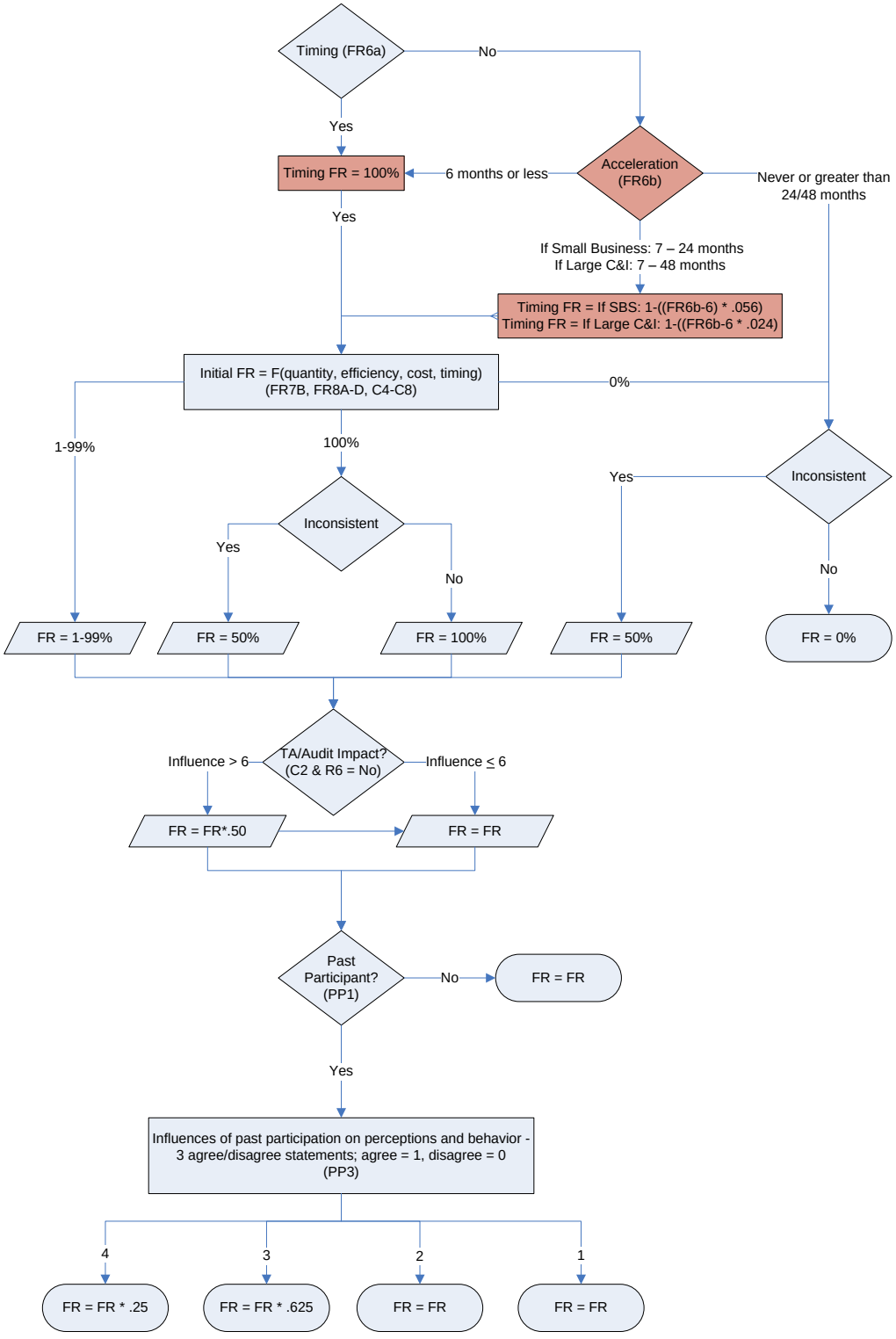


Figure F-2. Vendor Interview Trigger

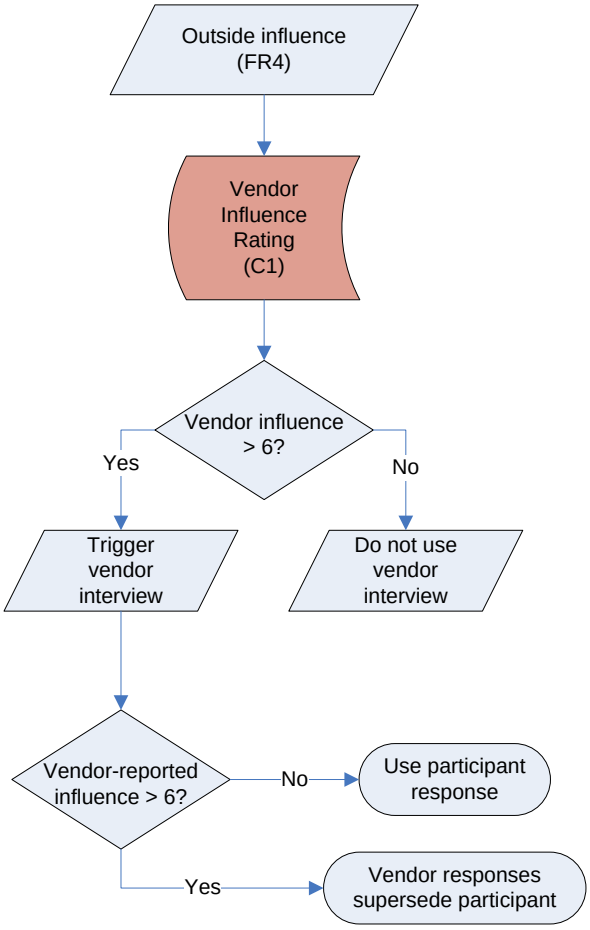


Figure F-3. Consistency Checks