

NYSERDA EMPOWER NEW YORKSM PROGRAM

Evaluation, Measurement, and Verification Plan

July 16, 2009

I. Introduction

The detailed evaluation plan presented in this document builds upon the brief evaluation plan included in NYSERDA's August 22, 2008 filing of the *System Benefits Charge Supplemental Revision for New York Energy \$martSM Programs (2008-2011)* for the EmPower New YorkSM (EmPower) Program. In order to revise and add detail to its original evaluation plan submittal, NYSERDA has incorporated feedback provided by the Department of Public Service (DPS) and the EEPS Evaluation Advisory Group (EAG), and has worked closely with its team of independent evaluation contractors to select the most appropriate evaluation approaches based on the current design of the program. This plan was developed to conform to the DPS evaluation guidelines released on August 7th, 2008 and to provide the highest level of rigor possible within the available resources.

As the EmPower Program ramps up to meet the aggressive EEPS goals, NYSERDA and its evaluation contractors will closely monitor aspects of that process such as participation levels, achievement of near-term goals, and other programmatic issues in order to adapt this plan, as needed, to provide the most relevant and useful evaluation. For example, adjustments may be needed to sample sizes or research issues if assumptions about the program do not develop as initially anticipated. As such, NYSERDA views this plan as a flexible, living document that will be updated, as necessary, with appropriate notice to DPS and other interested parties.

This evaluation plan was designed to constitute a comprehensive approach to assessing the entire EmPower Program which is supported by SBC and EEPS funding. NYSERDA will not differentiate between funding sources when conducting this evaluation effort.

II. Summary of Goals, Cost and Schedule for Evaluation Activities

The EmPower Program budget (3rd Quarter 2008 through 2011) consists of approximately \$23.6 million in EEPS funds and \$20.6 million in SBC funding, providing a total budget of \$44 million. The proposed evaluation budget is approximately \$859,000, which is less than 5% of total program funding.¹ However, NYSERDA believes this level of funding for evaluation is appropriate given the contribution of this program to the overall portfolio level savings, and the typically low variance in the residential population, as well as other factors. Evaluation budgets are detailed in Table 1.

¹ This evaluation budget includes only external contractor costs. Other overarching evaluation costs, including NYSERDA's internal evaluation management and statewide study costs, are additional; however, the total evaluation costs will not exceed 5% of program funding at the portfolio level.

Table 1. EmPower New York Evaluation Schedule and Budget

Evaluation Element	Estimated Budget and Completion						% of Total Evaluation Budget
	2009a	2010b	20110	2012	2013c	Total	
Impact Evaluation ¹	\$220,000	\$207,000	--	--	\$239,000	\$666,000	81%
Process and Market Evaluation ²	\$37,000	\$119,000		-	--	\$156,000	19%
Total	\$257,000	\$326,000	\$0	\$0	\$239,000	\$822,000	100%
<p>1. This plan describes a joint process and market evaluation for the EmPower Program. This plan was developed by NYSERDA's Process and Market Characterization and Assessment (MCA) evaluation contractors and includes aspects of both evaluation functions.</p> <p>2. Survey cost assumptions include a 20 minute survey of 600 EmPower program participants. Cost includes questionnaire review, pre-testing, CATI programming, advance letter, interview with head of household or spouse, coding, data processing with SAS, SPSS, Stata, and Excel files, preparation of Codebook, and documentation. Requires about a four week field period to get 60% response rate.</p> <p>a. The data collection portion of the 2009 Process and Market evaluation is estimated at \$10,000 and \$25,000 for the 2009 Impact evaluation.</p> <p>b. The data collection portion of the 2010 Process and Market evaluation is estimated at \$47,000 and \$26,000 for the 2010 Impact evaluation.</p> <p>c. The data collection portion of the 2013 Impact Evaluation is estimated at \$29,000.</p>							

The overarching goals of NYSERDA's **New York Energy \$martSM** and EEPS program evaluation efforts are to: (1) conduct credible and transparent evaluations, and (2) provide NYSERDA program staff and managers, the New York State Public Service Commission (PSC), Department of Public Service (DPS) staff, and other stakeholders with timely and unbiased information regarding program implementation. Specifically, the goals for the EmPower Program evaluation are to:

- (1) Establish rigorous and defensible estimates of the energy savings that can be attributed to the program;
- (2) Investigate coordination issues with New York utilities and weatherization agencies so as to improve program marketing, recruitment, and coordination of services to low income households;
- (3) Assess the market conditions related to the EmPower Program and to develop and provide suggestions to maximize the efficiency and effectiveness of the program;
- (4) Examine and document program activities and progress; and
- (5) Assess response to the program from the perspective of various stakeholders and customers; and make recommendations for program improvement.

III. EmPower Program Description and Goals

The focus of EmPower is on cost-effective electric reduction measures, particularly lighting and refrigerator replacements, as well as other cost-effective home performance measures (insulation, air-sealing, heating system repair and replacement, and health and safety measures). In-home energy use education provides customers with additional strategies for managing their energy costs. Participants are also invited to energy-use management and financial management workshops held in communities across the state.

Electric customers that live in 1-to-4 family homes or small multifamily buildings with 100 units or less, and either participate in a utility payment assistance program, or have a household income below 60% of State median, are eligible. There is no cost to the customer. In rental situations, measures that directly benefit the eligible tenant are eligible without a landlord contribution. Additional measures generally require a 25% landlord contribution. The energy efficiency services are delivered by a group of nearly 100 private contractors and Weatherization Agencies accredited by the Building Performance Institute (BPI).

The program will prioritize cost-effective efficiency measures for low-income households with high energy costs. The average annual energy savings for customers receiving electric reduction are estimated at 1,306 kWh and 28 MMBtu for those customers receiving home performance services. The program will supplement the efficiency services with energy use management and financial management education. It will develop an effective referral mechanism to target services to households with high energy burdens and improve coordination of complementary low income energy programs. The program will enhance the network of energy service providers, including private contractors and Weatherization Agencies.

Table 2 displays the MWh savings projected for both EEPS- and SBC-funded portions of the EmPower Program.

Table 2. Projected MWh Savings for EmPower Program (2008-2012)

	2008 (4 th quarter)	2009	2010	2011	2012	Total
EEPS MWh	0	8,015	10,686	10,686	0	29387
SBC MWh	3,985	5,725	2,482	2,482-	-	14,674
TOTAL	3,985	13,740	13,168	13,168	0	44,061

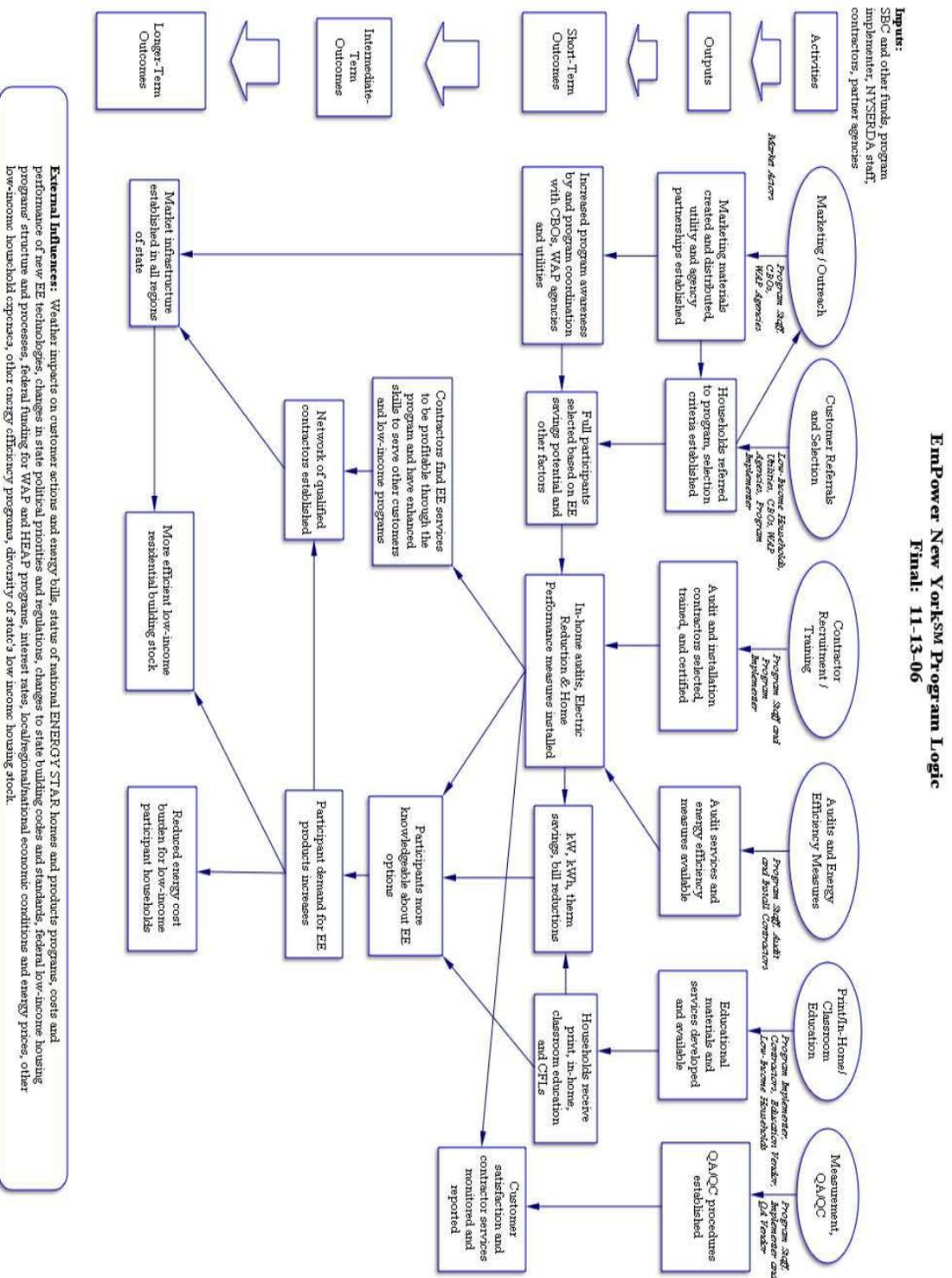
Source: NYSERDA, *System Benefits Charge Supplemental Revision for New York Energy \$martSM Programs (2008-2011)* As Amended, August 22, 2008 and revised March 12, 2009, Page 21, Table 5-1.

IV. Logic Model/Theory

Figure 1 presents the most recent logic model for the EmPower Program. As program evaluation efforts begin, a first step in the process will be to review the latest logic model and make updates to the model as necessary.

Logic modeling activities will occur early in the evaluation process after completion and approval of the Detailed Evaluation Plan. NYSERDA’s evaluation contractors convene logic model “workshops” with program staff to discuss program inputs, activities, outputs, outcomes, external influences and other elements that need to be documented in the logic model. The evaluation contractors then document these discussions in a brief program theory/logic report, which includes a logic model diagram for the program. NYSERDA will invite DPS Staff to participate in logic model workshops and review draft program theory/logic reports. Note that data collection plans of the Detailed Evaluation Plan might be modified somewhat in light of the resulting, program theory, logic model and performance indicators.

Figure 1. EmPower Logic Model



V. Market Characterization & Assessment Plan

The Market Characterization and Assessment evaluation will be conducted as a joint effort with the Process evaluation as described below. Please refer to that section for details on Market Characterization and Assessment activities.

VI. Impact Evaluation Plan

The EmPower Program (including the prior SBC-funded Weatherization Network Initiative) has served over 32,000 low-income households through September 30, 2008. Energy savings from participating households between July 1, 2006 and March 31, 2009 (SBC 3 time period) total 26.1 GWh Statewide, with 4.6 GWh (18%) occurring downstate in the Con Edison service territory.² EmPower offers a comprehensive audit, as well as energy education, electric reduction measures and home performance services to low income households in buildings with fewer than 100 housing units. EmPower has been expanded to include a wider range of measures designed to reduce natural gas consumption.

Impact evaluation for program years 2007 and 2008 is planned for 2009. In an effort to align the evaluation with the timing of expected savings, the first impact evaluation effort for the jointly SBC and EEPS-funded program is planned to take place during the latter half of 2010 for 2009 participants, and be repeated for 2010 and 2011 participants in 2013, assuming at least 12 months pre- and post-installation consumption data (24 months of data) are available.

This plan demonstrates a large improvement over the last Measurement and Verification analysis conducted on the program in 2007. The sample size has increased dramatically from site visits made to 20 projects to a billing analysis of all participants and a more focused billing analysis of a sample of approximately 600 households proposed in this plan. The increase in funding will also allow a pilot study of attribution (which had not been addressed by prior evaluations) and for impact evaluation to be conducted more frequently.

Research Objectives

The purpose of impact evaluation is to establish rigorous and defensible estimates of the savings that can be attributed to the EmPower Program. This process involves determining the realization rate for gross savings and the free-rider and spillover factors for net impacts. In both of these aspects of the impact evaluation, the evaluators need to determine how to achieve the desired precision, minimize the possibility of bias in the result and assess the validity of the results. Each of these key aspects of impact evaluation is discussed briefly below.

Determine Realization Rates for Gross Savings

A critical component of the impact evaluation is to develop rigorous estimates of the realization rates for gross electricity, demand and other fuel savings. There are numerous approaches to achieve this goal, including billing analysis and verifying the installation and the estimation of savings for a representative sample of program participants. An optimal strategy is proposed herein based upon

² NYSERDA, **New York Energy \$martSM** Program Evaluation and Status Report, Quarter Ending March 31, 2009, Report to the Public Service Commission, May 2009, Table 4-1.

level of effort required, availability of data and cost. However, a secondary strategy, and the data required should that alternative approach need to be employed, is also detailed below.

Attribution

Non-low-income programs require an assessment to determine the extent to which impacts are program-induced (rather than naturally occurring). This objective is often accomplished through estimation of the ratio of impacts for those that would have taken the actions without the program (free-riders) compared to program savings and the ratio of the savings from actions taken outside NYSERDA programs but due to the program (spillover). The combination of these components in the form of a net-to-gross (NTG) ratio becomes the adjustment factor to derive net impacts.

Low-income programs are often assumed to have no free-riders or spillover based on the belief that occupants cannot afford to take any of these actions without the program. This assumption will be tested through a pilot project to assess a net-to-gross (NTG) ratio and determine if NTG analyses should be pursued with more rigor in future EmPower Program evaluation efforts.

Precision and Bias

Sample sizes will be designed to target 90/10 precision for the program savings overall statewide.³ A billing analysis of all participants with sufficient billing data would also allow for the estimation of savings at the utility level with no sampling, thus eliminating concerns about sampling precision. The Impact Evaluation Team will also field a telephone survey on this smaller sample of billing analysis participants to address the potential bias created by external events that affect energy consumption but are not in any way related to program activity. Consequently, the feasibility of estimating savings at the utility level through a model that incorporates telephone survey respondents will need to be further considered.

Methods will be selected to minimize self-selection, non-response and other sources of bias, to the extent possible. Bias mitigation through telephone surveys is mentioned above and covered in greater detail under the billing analysis section below. Another example is that the non-response rate for telephone surveys can be reduced by ensuring that several attempts are made to contact each potential respondent at different times of the day.

Activities

Gross Impacts

Billing Analysis

One of the most reliable impact evaluation methods for energy efficiency programs targeting existing buildings is conducting a billing analysis using pre- and post-energy use data to statistically analyze average energy savings. Billing analysis is feasible when savings are measurable in comparison to total household consumption (often recommended to be 8 to 10% of pre-installation consumption),

³ The sample size depends on the type of statistical analysis being conducted and the type and variability of the specific parameters to be estimated. For example, a simple random sample required to achieve 90% confidence and 10% sampling precision for a yes/no question is about 67 for a large population. However, if the variable of interest is the realization rate and the coefficient of variation is 0.75, a simple random sample would require a sample size of 152 to achieve the same precision and confidence level.

when the population is fairly homogenous and when the program tracking system provides sufficiently detailed information to support the analysis. It is possible to quantify lesser savings when the sample is highly homogenous or the sample size is quite large.

In addition to the improved rigor, billing analysis is generally less costly than other approaches due to the fact that the billing data is already collected by the utilities. Consequently, most of the population can be included in the model with little additional cost required for data collection. A related advantage of using billing analysis on all participants is that reliable estimates at the utility level are quite likely. Few other rigorous impact methods can allow utility level estimates for this statewide program with only a small incremental cost.

In one important respect, the EmPower Program is a good fit for a billing analysis. This program targets average annual savings of approximately 1,300 kWh per household, or about 20% of use assuming average annual household electricity use of 6,000 kWh. In New York homes are predominantly heated by natural gas, comprising 47% of fuel usage by the residential sector.⁴ For natural-gas heated homes, insulation, weatherization and heating equipment measures would be significant enough in this cold-weather state to obtain reliable savings estimates from billing analysis. This preliminary review suggests that the household savings levels should be large enough to estimate through billing analysis.⁵

Another critical aspect of the program evaluation is that the low income sector can be difficult to evaluate for a number of reasons. Previous experience with low income evaluations in California indicates that a substantial part of the low income population tends to be more mobile. In addition, low income households tend to use less energy than the residential market as a whole and may be more likely to have disruptions in their utility service.⁶ All of these factors tend to add a higher degree of uncertainty to impact evaluations of low income programs.

The Impact Evaluation Team has not yet had the opportunity to conduct a thorough review of the EmPower program-level data. An initial review reveals that many of the fields necessary for billing analysis are tracked by program staff, although the available data may need to be supplemented through a participant survey.

EmPower Participants

Residents of both single family and multifamily (up to 100 housing units) dwellings are eligible to participate in EmPower. Conducting an aggregated billing analysis across all or most program participants requires a consistency in the observational unit, *i.e.*, the home. Participants living in master-metered buildings will have to be omitted from the analysis due to the disparity in size between a single family home and a master-metered multifamily building that could contain up to 100

⁴ *Patterns and Trends*, New York State Energy Profiles: 1992-2006, published January 2008 by NYSERDA.

⁵ It will not be possible to conduct a billing analysis for the heating-related measures for homes with an oil or propane primary heating system due to the complexity of obtaining and interpreting the billing and delivery/storage records. Given the similarity in the analysis of heating-related loads, the realization rates for the heat-related measures from the natural gas analysis will be applied to the savings estimates for oil and propane heated homes. This strategy is based on the assumption that the accuracy (level of bias) of the algorithms used by the program for estimating oil and propane savings is the same as those applied by the program for natural gas heated homes.

⁶ "Impact Evaluation of the 2005 California Low Income Energy Efficiency (LIEE) Program," prepared for SCE, PG&E, SDG&E and Southern California Gas by West Hill Energy and Computing, Inc, August, 2008. See Chapter 4.

units.⁷ The Impact Evaluation Team understands that few EmPower participants live in master-metered units, suggesting that this omission should not introduce a substantial source of bias to the analysis. Upon further review of the program-level data, the Impact Evaluation Team will assess whether master-metered multifamily units are a sufficiently large component of the program to require an alternative approach, such as selecting a sample of these participants for a per-building billing analysis, *i.e.*, IPMVP Option C.

Billing Analysis Issues

In order to conduct this analysis, NYSERDA will require utility account numbers, electric and natural gas consumption records (kWh, kW, therms and interval/advanced meter data) for participants and non-participants (if possible), to be provided electronically in a readily accessible format. NYSERDA currently obtains account information and permission from the participant to obtain energy use data from the utility or utilities. Non-participant billing data may be more difficult to obtain, and the Impact Evaluation Team will work the NYSERDA and the utilities to pursue this approach. The Impact Evaluation Team recognizes the importance of protecting confidentiality of the consumer's data (see *Customer Data Guidelines*, June 12, 2009).

A challenge in conducting a residential billing analysis is that residential use tends to vary tremendously from one home to the next depending on the number of occupants in the home, the types of energy end uses, life style and numerous other non-program related factors. The Impact Evaluation Team will use a regression technique based on Analysis of Covariance (ANCOVA) to estimate program savings by allowing each participant to act as his or her own control and accounting for the energy characteristics of the home and other effects (such as weather) within the statistical analysis. If sample size is sufficient and the regression estimators are significant, savings will be reported for each group of related measures (such as air sealing, hot water conservation, lighting, etc.).⁸

The Impact Evaluation Team will carefully review the data and assess the results to ensure that the savings estimates are statistically sound. Testing for violation of statistical assumptions will be conducted and any violations will be mitigated to the extent possible. The information-theoretic approach to model selection will be employed to ensure that the selection of the final model is based on objective statistical standards.⁹

Non-Program-Related Impacts on Energy Use

External factors often have as much, or more, impact on energy use than efficiency programs. These external impacts can introduce either an upward or downward bias to the results. For example, a

⁷ As long as the sample size is sufficient, it would be possible to conduct a billing analysis using the single-family homes and individually-metered multifamily units.

⁸ In some cases, measures are effective when they are installed as a package rather than individually. Air sealing measures are a good example. It would not be possible to estimate savings separately for weatherstripping, caulking, sealing bypasses, etc. However, these measures can be treated as a group and it may well be possible to estimate savings on that basis.

⁹ In billing analysis, the analyst makes many decisions regarding the statistical characteristics of the model and the specific parameters to be included. Thus, there are typically a number of possible models that could be used to estimate savings. The information-theoretic approach provides an objective framework for selecting the best model among a series of competing candidate models. Please refer to *Model Selection and Multimodel Inference* by Kenneth Burnham and David Anderson, Springer-Verlag, NY, 2002.

billing analysis may indicate that participants on average are saving 100 kWh per year, but it may be that on average residential electric consumption increased by 100 kWh per year during the same period, suggesting that the actual program impacts are 200 kWh per year. Given that the national economy appears to be moving into a period of contraction, one would expect that energy consumption may be reduced across the board, making it likely that program savings may be under- or over-estimated, depending upon the timing of the pre- and post-installation periods.

While the ANCOVA method controls for the characteristics of the home that are stable over time, it is possible that the estimation of program impacts can be affected by other factors that do change over time. These types of changes can be conceptualized in two broad categories:

- changes in the overall economy that affect the residential market in a global way, such as volatile gasoline prices, unemployment rates, or an increase in home heating costs.
- individual changes that affect specific homes, such as acquiring new household members, taking a longer vacation, or having a change in one's work schedule

These issues need to be considered and addressed separately.

Addressing Global Economic Impacts

There are two common approaches to address the global factors within the statistical analysis: 1) include a non-participant comparison group directly in the billing analysis, and 2) incorporate trends lines into the analysis. Unfortunately, the first strategy can introduce net effects into what should be an estimate of gross savings. In the end, a billing analysis that includes both participants and a non-participant comparison group will likely produce savings estimates that are somewhere in between net and gross effects and, thus, difficult to interpret with any degree of accuracy.

Incorporating trend lines will be the preferred strategy applied in this evaluation. The Impact Evaluation Team will rely on third party economic data, often available through public sources, and non-participant consumption data to develop these trend lines.

Constructing a representative comparison group for the purpose of developing a kWh consumption trend variable requires a reliable method of identifying low income residential customers. One common approach is to use program participants in the next program year and analyze their billing data in the period prior to program participation. For example, if the impact evaluation for program years 2007 and 2008 is conducted in 2010, then participants who completed installations prior to July of 2008 could be included in the billing analysis, and the comparison group could consist of participants who enrolled in July of 2009 or later. Another possibility is requesting assistance from the New York utilities to identify their low income customers, either through their rate structures or other means.

Non-participant information is a critical component to understanding and addressing external trends in energy use. The Impact Evaluation Team will carefully review all options with the goal of constructing a reliable non-participant comparison group.

Addressing Within-Home Factors Affecting Energy Use

The Impact Evaluation Team also intends to field a participant survey to inform the billing analysis and to assess whether more detail on within-home changes will effectively improve the model. This

survey will be designed to obtain additional information regarding typical changes occurring with the residence during the pre- and post-installation period (such as replacing a refrigerator, and changes in schedules and occupancy). In addition, this survey also provides the opportunity to investigate the more personal impacts of the national and regional economic trends, such as changes in employment status. Since this type of information may be sensitive to some participants, the Impact Evaluation Team will consider how to incorporate this line of questioning into the survey and how the question can be worded to minimize non-response. The specifics of the survey will be developed as part of the work plan.

Given the wide variability in residential energy use, which is often amplified in the low income sector, a large sample size will be necessary. The billing analysis will be conducted with this subset of participants who receive the phone survey and with all participants with complete billing data. The results of these two analyses will be compared for reliability.

Previous experience in this type of evaluation suggests that the approach of collecting additional information on patterns in individual homes does not always result in a better specified model.¹⁰ Conducting a phone survey necessarily restricts the sample size, and in some cases the sheer volume of including all participants with complete billing records provides more reliable results. This situation is more likely to occur when the program savings are small in comparison to total energy consumption.

The Impact Evaluation Team has concluded that, given EmPower's high estimated savings per home, adding information from a telephone survey to the billing model has a good chance of producing statistically significant and unbiased estimates of gross impacts. Even if the smaller sample size does not improve the model, the phone survey will still provide critical information about the external influences on energy use for this variable market segment.

Demand Savings

Demand (kW) savings will be estimated by applying a kWh/KW factor to the energy savings. For example, the most recent kWh/kW ratios per measure in the Deemed Savings Database (DSD) could be used to derive the program's kW estimates. This approach is commonly used in the residential sector, and improved rigor for demand savings generally entails substantial increases in budget.

Coordination with Other Evaluation Efforts

The federal Weatherization Assistance Program (WAP) may also be fielding an impact evaluation of the WAP activities in New York with the Department of State (DOS), and the potential exists to coordinate with this effort to obtain billing records from the utilities and to conduct other aspects of the EmPower impact evaluation jointly with the DOS. This possibility will be explored to ensure that NYSERDA does not lose the opportunity to collaborate with the federal and state evaluation efforts.

¹⁰ Xenergy (now KEMA) conducted an impact analysis of the Low Income Energy Efficiency program in California for program year 2000 which included a large-scale billing analysis and a telephone survey of 1,000 participants. Xenergy ultimately concluded that the results from large scale billing analysis were more reliable than the smaller scale analysis that included only the telephone survey respondents. The scale of the savings were in the range of 5% of total energy use, and the extremely large sample size allowed for the estimation of these small effects. This study is available on the CALMAC website.

Alternatives to the Billing Analysis Approach

If the consumption data are not forthcoming, the Impact Evaluation Team would aim to complete an adequate number of site visits with on-site measurements to meet 90/10 confidence and precision for the statewide program. This approach would require pre/post measurement or detailed site visits on a sample of homes and delaying services to customers until after the pre-period metering was complete, which may be particularly problematic for this program as it is targeted to the low-income sector. Furthermore, NYSERDA may need to offer financial incentives to help reduce this significant negative impact to customers.

If this fall-back option must be implemented, the Impact Evaluation Team will attempt to make direct measurements and use the most rigorous impact evaluation method that can be obtained within budget given the inability to do large-scale billing analysis. This approach would also require reallocation of the evaluation budget. Another consequence would be the inability to develop rigorous savings estimates at the utility level without a substantial increase in the budget.

Attribution

EmPower provides services to a population with limited means of purchasing energy-efficiency goods and services on their own. Due to an anticipation of low free-ridership and spillover rates, the SBC-funded EmPower Program has not been examined for attribution of program impacts. Many jurisdictions (including Vermont and California) simply assume that low-income programs do not have any free-riders or spillover. In addition, spillover is not generally associated with resource acquisition-style retrofit programs such as EmPower.

Given that this aspect of the program has not been evaluated in previous years and there is no empirical estimate of freeriders for this program, the Impact Evaluation Team is planning to conduct an initial pilot study by incorporating questions about net effects into the telephone survey to be fielded as part of the gross impact evaluation. This approach is intended to be a low cost effort to assess whether there may be free-ridership or participant spillover associated with this program. Consequently, relying on self-reporting through the telephone survey is a reasonable first step. If this initial pilot study demonstrates some degree of free-ridership or spillover, then the attribution component of the subsequent impact evaluations may be expanded, as needed.

The Impact Evaluation Team also considered the possibility of using a comparison group to assess net effects; however the costs associated with the additional surveys required for such a study did not seem warranted since there is no reason to assume that free-ridership or spillover will be significant for this program. The NTG approach will be re-evaluated based on the results of the initial pilot study.

Populations/Samples

The sampling approach for use in the estimation of gross impacts and in the pilot NTG projects is discussed in more detail below.

Gross Impact Sampling

For the verification of gross savings, the preferred impact evaluation approach is a billing analysis. In this case, all participants with sufficient billing history will be included in the analysis and no

sampling will be necessary. Thus, there will be no sampling error, and it will also be possible to conduct the analysis at the utility level.

The Impact Evaluation Team also intends to field a telephone survey to expand the data inputs into the billing analysis. This survey will be leveraged to support the net-to-gross pilot study. There is no clear consensus on how to determine sample sizes to reach a specific precision/confidence level for the type of complex, multivariate analysis planned for this program. Based on previous experience with other low-income program evaluations and given the expected savings in comparison to pre-installation energy use (20%), the Impact Evaluation Team estimates that 600 completed surveys are likely to be a sufficient sample size to estimate household savings at the target relative precision of 90/10 upstate/downstate.¹¹

The models will be run including all participants with sufficient billing records and also with the restricted sample of 600 from the telephone surveys. The results from both analyses will be presented and compared, with a discussion of the reliability, precision and potential bias (if any). If the precision target cannot be met with the restricted sample, either at the program or upstate/downstate level, the results from the more comprehensive model can be used as the source of final program savings.

A stratified random sample for the telephone survey will be selected based on annual household energy use and region, for the purposes of ensuring that the sample represents the total program population in terms of these key parameters.¹² The utility billing records will be needed prior to the sampling to ensure that sufficient billing history is available for all participants selected for the telephone survey.

Should site visits be required, the Impact Evaluation Team will attempt to complete a sufficient number of site visits with metering to estimate savings with the 90/10 confidence and precision standard statewide as well as on an upstate versus downstate regional basis. However, it should be noted that this approach is sub-optimal and may have negative impacts on program participation. Additionally, using this evaluation approach would require funding from elsewhere within NYSERDA's evaluation budget to support the substantial costs of on-site surveys with the sample size required to attain 90/10 confidence and precision at the program level. If this evaluation approach is required due to the inability to obtain a comprehensive set of billing data, then site visit data will be collected and analyzed by the Impact Evaluation Team following established evaluation protocols.

Attribution Sampling

The attribution questions will be incorporated into the Impact Evaluation Team's telephone survey of participating homeowners and renters. Given that the sample size of 600 was established to provide a sufficient sample for a complex multivariate regression model, it is more than adequate to estimate the attribution parameters of interest (such as the likelihood of free-ridership) at the 90% confidence and 10% precision standard statewide, as well as on an upstate/downstate regional basis. This sample

¹¹ The power analysis was conducted with conservative assumptions, and indicated that 540 participants would be required. Given previous experience with low-income residential billing analysis, the Impact Evaluation Team further increased the sample size to 600.

¹² This sample size might not produce statistically significant results separately for each region or each category of pre-installation consumption.

size may also produce reliable utility level results that meet the 90/10 confidence/precision standard (depending upon the sampling requirements for the billing analysis versus the attribution study).

Evaluation Issues

There are a number of issues that may complicate the sampling and other aspects of the evaluation, as listed below.

- Utility usage data will be needed for all participants and possibly non-participants (for constructing trend lines). This approach will require close cooperation with the utilities to obtain billing history for all participants. Even if all of the participants are correctly identified by NYSERDA and matched by the utilities, it is likely that only half or less of the participants will have the full two years' worth of billing records required for the analysis.¹³ In addition, utilities archive billing data at regular intervals and the NYSERDA data request will need to be made before data from the pre-installation period would be archived.
- Aggregated billing analysis will only work with similar observational units, and single family and master-metered multifamily buildings cannot be combined in the same model. The incidence of participants living in master-metered buildings will have to be assessed to determine the impacts on the number and characteristics of the EmPower participants who can be included in the billing analysis model for single-family/individually-metered housing and the possibility of developing a billing analysis model for master-metered buildings, if needed.
- Incorporating trend lines into the billing analysis is one method to ensure that general market trends are taken into account. Possibly the most important trend in this context is the change in energy consumption. The preferred approach to developing these trend lines is to obtain billing data from the utilities for future program participants or for the general population that best matches program participants. The Impact Evaluation Team will need to work with NYSERDA and the utilities to determine whether there is a simple mechanism for identifying low-income customers.
- The impacts of attrition will need to be assessed to ensure that it does not introduce a systematic bias into the results. One source of attrition is due to mobility *i.e.*, participants may not stay in the same location long enough to accumulate two year's worth of billing history. Erratic billing records or many estimated reads is also sufficient reason to remove a home from the billing analysis. The characteristics of the final group of participants left in the analysis will need to be compared to the total program population to assess the possibility that some particular group of participants may be omitted from, or severely underrepresented in, the analysis.

These issues will need to be resolved to ensure that the sampling can proceed within the required time frame.

Data Collection

To be able to conduct the sampling and proceed with the evaluation, the Impact Evaluation Team will need the following information, at a minimum, from NYSERDA's EmPower staff:

¹³ In general, low-income populations tend to be more mobile than the general population.

- Project level information, including address, contact information for the homeowner or renter, contractor identification (and a list of contractor contact information), entity (contractor/WAP agency) providing the services
- Measure level information, such as a description of the measure, quantity installed, the energy savings (electric, gas and other fuels), demand savings, measure life, installation costs
- House level information, including the size of the house, the number of occupants, the fuels used for space and water heater, other major electric and gas end uses

In addition, critical information will need to be collected from utilities, as described below.

- Utility consumption data for participants (both electric and natural gas), covering the date of the read, account number, premise number, amount of energy used, tariff, rate class, whether the read was estimated or actual, city or zip code, and (if available) weather station
- Utility consumption data for future program participants or for the general population that best matches program participants (same fields as listed above for participant utility consumption data for use in developing trend lines)
- Weather data, which may be available from the utilities or from the national weather service (Source: National Oceanic and Atmospheric Administration (NOAA))
- Economic trends, such as gasoline prices and unemployment rates, often available from public sources, such as the Bureau of Labor Statistics and U.S. Department of Energy's Energy Information Administration

Primary data collection planned for the 2009 impact evaluation is shown in Table 3.

Table 3. EmPower Program Impact 2009 Evaluation Survey Specifics

Target Group	Estimated Population Size	Estimated Sample Size	Expected Sampling Precision	Survey Administration By	Expected Start of Fielding
Participating Homeowners/Renters - Gross Savings and Attribution	~3,200/six months	600	90/10	Survey Contractor	Fall 2009

The impact evaluation schedule and budget is presented in Table 4. Most of the costs for the proposed impact evaluation are the costs of preparing utility bills for analysis, the billing analysis itself, and the telephone survey of participants. These items constitute approximately two-thirds of the total costs for the proposed impact evaluation. The fixed costs constitute the other third of the total costs and cover the costs for project management, instrument development, further sample/population analyses, status updates, and reporting.

For gross savings, the level of the analysis as presented for 2009 is expected to be repeated in 2010 and 2013. Thus, in the later years, if the smaller analysis using 600 participants proves superior to the larger analysis using all 2009 participants with sufficient billing data, a sample of approximately 600 participants will be selected for the telephone survey and a billing analysis for all participants with sufficient billing history will be undertaken. In addition, the pilot attribution component of the evaluation will be assessed and re-designed based on the lessons learned in the 2009 evaluation.

The post 2009 budgets are based upon the 2009 methods, assumptions and costs. Fixed costs such as those associated with design, reporting, and management include both escalation and significant discounts, accounting for the likelihood that replicating a study is less expensive than executing it for the first time.

Table 4. EmPower Program Impact Evaluation Schedule and Budget

Evaluation Element	Estimated Budget and Completion					
	2009	2010	2011	2012	2013	Total
Impact Evaluation ¹	\$220,000a	\$207,000b	--	--	\$239,000c	\$666,000d

¹ Survey cost assumptions include a 20 minute survey length on 600 EmPower program participants. Cost includes questionnaire review, pre-testing, CATI programming, advance letter, interview with head of household or spouse, coding, data processing with SAS, SPSS, Stata, and Excel files, preparation of Codebook, and documentation. Requires about a four week field period to get 60% response rate.

a The data collection portion of the 2009 Impact evaluation is estimated at \$25,000.

b The data collection portion of the 2010 Impact evaluation is estimated at \$26,000.

c The data collection portion of the 2013 Impact Evaluation is estimated at \$29,000.

d This includes SBC III and EEPS evaluation funding. The 2009 impact evaluation is for the 2007-2008 EmPower Program, counting towards New York’s 15X15 goals.

VII. Process and Market Evaluation Plan¹⁴

The process evaluation of the EmPower Program is intended to be a formative investigation of the functioning of the program in order to identify areas of improvement. It also includes research to better define the market and context within which the EmPower Program operates as well as to define changes in the market over time with a specific focus on market indicators that are likely to be impacted by program offerings.

Two previous process evaluations have addressed overall program operations and issues related to referrals, recruitment, program expansion capability, participant satisfaction, location and effectiveness of the energy efficiency workshops, the effectiveness of the quality control/assurance systems, and the process of working with weatherization contractors and agencies. These previous evaluations were spread out over a four year period due to limitations in funding.¹⁵ The planned evaluations will build on these earlier efforts and provide feedback on how the program processes are functioning with the additional EEPS funding. The process evaluation will also include a market

¹⁴ As noted in the *System Benefits Charge Supplemental Revision for New York Energy SmartSM Programs (2008 – 2011)*, As Amended, August 22, 2008, NYSERDA will conduct a joint process and market evaluation for the EmPower Program. This plan was developed jointly by NYSERDA’s Process and Market Characterization and Assessment (MCA) evaluation contractors and includes aspects of both evaluation functions. A separate MCA evaluation plan will not be prepared.

¹⁵ This evaluation plan builds on past process evaluation findings integrated into the program over time including findings from the Process evaluation report from July 2007 and the present report published in March 2009. Recommendations from these reports are presented to program staff, who provide commentary in the August Quarterly and Annual Energy Smart evaluation reports. This commentary includes how staff specifically implemented suggested findings from the evaluation contractor into the program design and operation.

assessment using geographical information system (GIS) analysis to match the availability and use of program services relative to the identified low income population.

The process/market evaluation of the combined SBC/EEPS program will begin in late 2009 and be completed in 2010. Should program and evaluation funding be continued beyond 2011, another follow-up process/market evaluation could begin in 2011 and be completed in 2012 to track progress in the market and identify any additional program improvement opportunities; a research plan will be developed prior to that time to reflect the specific formative evaluation issues of concern.

Research Objectives

The primary process evaluation objectives are noted below. These objectives reflect issues specific to the expanded program efforts as well as follow-up on issues identified in previous evaluations, such as recommended improvements to the workshops and strategies for reaching referred customers. In order for the process evaluation to provide the greatest value, other relevant or necessary objectives may be added as the timing of this research draws closer.

1. Update the EmPower Program Logic Model Report¹⁶ to ensure the document accurately reflects the current program design and state of the market
 - a) Research the designs and implementation schedules of complementary energy efficiency programs fielded by utilities and third-party administrators to identify potential leveraging opportunities and coordination issues.
 - b) Prioritize measurement indicators and researchable issues.
 - c) Generate feedback regarding proposed program design and implementation strategies and suggest opportunities for program improvement, if any are observed.
2. Investigate coordination issues with New York utilities and weatherization agencies so as to improve program marketing, recruitment, and coordination of services to low income households.
 - a) Assess whether referral letters are being read and understood by customers.
 - b) Assess whether utilities are identifying potential participants in a consistent manner and coordinating their efforts with NYSERDA.
3. Assess market conditions relevant to the EmPower Program and develop and provide suggestions to maximize the efficiency and effectiveness of the program.
 - a) Assess the geographic distribution of referrals, applications, services and workshops.
 - b) Assess the distribution and relative experience of contractors and agencies.
 - c) Estimate the population of households eligible to participate in the Program and assess program penetration into this population (along spatial and temporal dimensions).
 - d) Examine market share of the nearly 100 BPI-accredited private contractors and Weatherization Agencies currently delivering program services and identify additional market opportunities.

¹⁶ NYSERDA, *EmPower New YorkSM Program – Program Logic Model Report*, November, 2006.

4. Document program activities and progress and assess response to the program from the perspectives of program staff, utilities, agencies, implementation contractors, weatherization building professionals, and residents of low income households.
 - a) Assess satisfaction with program processes and services from each groups' perspective.
 - b) Assess barriers to program participation for low income households and for agencies and contractors.
 - c) Assess how outreach and coordination with utilities, agencies and contractors can be enhanced to address barriers to participation.

Activities

The process evaluation team will conduct interviews with NYSERDA and program implementation staff, utilities staff, agency representatives, and weatherization contractors as well as surveys with participating and with referred nonparticipating¹⁷ low-income households and nonparticipating BPI certified contractors.¹⁸ The interviews with program and implementation staff will examine how the program operates and what is and is not working well from their perspective. The interviews with participating utility and weatherization agency representatives, and contractors as well as the surveys with nonparticipating weatherization contractors will identify implementation issues, program barriers, and opportunities to enhance the program.

The surveys with participating and referred nonparticipating low-income households will address program process issues such as awareness, experiences and satisfaction with the program and marketing and recruitment approaches, and ask questions to identify barriers that might restrict program participation. In addition, the surveys will include market characterization and assessment questions such as bill payment history, previous program participation, average tenure of potential participants at their current address, and access to various sources of energy information. Research findings generated by the evaluation will be related to the outputs and outcomes anticipated by the program logic model.

A GIS analysis will examine the utility, contractor and agency proximity, housing stock and income data from the census and then plot the location of completed projects, and the location of all referrals against the population of general low-income households. This analysis will be used to address Objective 3 and will help determine how well the distribution of services matches the distribution of eligible low-income households.

¹⁷ EmPower relies on utility referrals. 'Referred nonparticipants' distinguishes customers who are referred to the program by the utilities and do not participate from nonparticipating income-qualified customers who have not been referred to the program by the utilities.

¹⁸ The evaluation team will explore opportunities to aggregate primary data collection efforts across programs into sector-wide or market-wide efforts. Doing so may help: 1) avoid duplication of effort in interviewing sets of market actors common to many programs (*e.g.*, weatherization contractors), and 2) hedge against the risk of overlooking certain market sectors not explicitly targeted by specific program offerings. In addition, the evaluation team will remain aware of the activities of the EAG's subcommittee on statewide studies to again avoid potential duplication of effort but also to determine how best to supplement any statewide studies approved by the DPS. Results of these efforts will be discussed in the final project workplan.

Populations/Samples

Table 5 below shows recommended populations and sample sizes. NYSERDA's process evaluation contractor team, currently led by Research Into Action, will conduct interviews with NYSERDA program staff and implementation contractor staff, referring agencies and with weatherization contractors, and with New York electric and gas utility representatives and DPS staff. Surveys will be conducted with nonparticipating weatherization contractors, provided such a list can be developed.¹⁹ Surveys with low-income households will be conducted using lists maintained by the EmPower Program contractor; these include lists of participating and referred nonparticipants. These lists will be stratified by utility service territory to ensure qualified customers of each utility are represented

Data Collection

The specific contractor team that will lead each data collection effort and the expected start date of fielding is also shown in Table 5. Data collection will begin in late summer 2009 with interviews with NYSERDA and program contractor staff; these interviews will each last about one hour. At that time the process team will begin working with program staff to assemble lists of agencies and weatherization contractors working in the program, as well as contractors not working in the program. In September and October 2009, the process team will conduct interviews with participating agencies and weatherization contractors, with New York electric and gas utility representatives and with DPS staff; these interviews will last 30-45 minutes. The surveys for participants and referred non-participants and for nonparticipating weatherization contractors will be developed in November and December 2009 and implemented between January and March 2010. The surveys of participants will last about 20 minutes, those with referred nonparticipants will last about 10-15 minutes, and those with nonparticipating weatherization agencies and contractors will last about 10 minutes. Non-response bias is typically addressed by repeated calls during varied times of day and is not anticipated to be a concern for these populations. Analysis and report writing will be completed by June 2010.

¹⁹ The list development will begin with Building Performance Institute contractors not participating in EmPower as well as list of weatherization contractor association members not participating. Possible challenges in identifying samples for these groups may arise if these lists cannot be obtained.

Table 5. EmPower Process and Market Evaluation Survey Specifics

Target Group	Estimated Population Size	Estimated Sample Size	Expected Sampling Precision	Administration By	Expected Start of Fielding
Staff and Implementation Contractor Interviews	12a	8	NA	Process Team	August 2009
Utilities & DPS	8	8	NA	Process Team	Sept 2009
Referring Agencies	98	15	NA	Process Team	Sept 2009
Participating Weatherization Providers	88b	38	90/10c	Process Team	Sept 2009
Nonparticipating Weatherization Providers	150+	68	90/10d	Survey Team	Jan 2010
Participating Households	~15,000	150	>90/10d	Survey Team	Feb 2010
Referred but Non-Participating Households	25,000	150	>90/10d	Survey Team	Feb 2010

a. Key staff includes 4 NYSERDA, 4 CSG and 4 Honeywell staff.
 b. These include 7 Vendors, 44 agencies currently active (including those under AEA) and 37 private contractors.
 c. Assumes proportional sampling, 2-tailed test, finite population correction, absolute precision
 d. Assumes proportional sampling, 2-tailed test, absolute precision

Schedule and Evaluation Budget

Table 6 presents the evaluation budget and schedule for the EmPower process and market evaluation.

Table 6. EmPower Program Process and Market Evaluation Schedule and Budget

Evaluation Element	Estimated Budget and Completion					
	2009	2010	2011	2012	2013	Total
Process and Market Evaluation	\$37,000a	\$119,000b	--	--	--	\$156,000

a. The data collection portion of the 2009 Process evaluation is estimated at \$10,000.
 b. The data collection portion of the 2010 Process evaluation is estimated at \$47,000.

VIII. NYSERDA Evaluation Process

This evaluation plan is an early, but important step in NYSERDA’s evaluation planning and implementation process. It is NYSERDA’s understanding that DPS Staff wish to be involved as a reviewer/participant in the following parts of the evaluation process: detailed evaluation plans, project kick-off meetings, workplans (including sampling, statistics and modeling issues), data collection instruments, interim results reports (as applicable), presentation of evaluation results, and draft evaluation reports. NYSERDA will conduct evaluation planning and implementation in an open and transparent manner, and will invite DPS Staff participation in the designated aspects of the process

and any others upon DPS' request.²⁰ Should DPS Staff choose to modify the level or manner of their involvement, NYSERDA should be notified about the change(s). DPS Staff should also choose when and how to involve their evaluation advisor consultant team in NYSERDA's evaluation processes, should directly provide any materials and information necessary for their advisor consultant team to fulfill this role, and should notify NYSERDA about the type and level of advisor consultant involvement.

An important goal of NYSERDA's evaluation effort is to provide early feedback to program staff to help inform and improve program implementation. NYSERDA accomplishes this goal in several ways:

1. Ongoing communications between the NYSERDA evaluation staff and evaluation contractors to identify issues that need to be brought to the attention of NYSERDA program staff, DPS Staff, and other involved parties.
2. Interim results reports may be generated, sometimes at the request of NYSERDA program staff and sometimes by initiative of NYSERDA's evaluation team and contractors, where early results are required or deemed useful prior to completion of the full evaluation effort.
3. Presentations of draft evaluation results held with NYSERDA evaluation contractors, evaluation team, program staff, and DPS Staff before evaluation reports are written provide feedback on the programs as soon as possible, and provide evaluation contractors with additional perspective and context that will be useful in reporting final recommendations.

Upon completion of final evaluation reports, the NYSERDA evaluation team will also provide support and assistance to program staff with regard to implementation of recommendations and program improvements.

IX. Reporting

Final reports will align with requirements set forth in the DPS evaluation guidelines, and will include: methodology, key results, recommendations, summary and conclusions, and appendices with detailed documentation.

Upon completion of each major evaluation study effort, findings and results will be communicated by NYSERDA's evaluation contractors and evaluation staff to NYSERDA program staff. Actionable recommendations and information on program progress toward goals will be provided as input to the program design and improvement process. NYSERDA's evaluation staff will follow up regularly with program staff on recommendations arising from the evaluation and the status of their consideration or adoption of these recommendations.

²⁰ In order to maintain transparency, and allow for confirmation checking and follow-up analysis, evaluation data will be maintained by NYSERDA and made available to DPS on an as-needed basis. NYSERDA will continue to maintain its secure "data warehouse" which includes data files, code books, and analysis files which can be made available in electronic form to DPS upon request. In order to provide a comprehensive record of each study conducted, the data warehouse also holds copies of final evaluation reports and appendices, including blank survey instruments, although these documents will be made available to DPS and publicly upon completion of each evaluation project.

NYSERDA's evaluation staff will prepare quarterly and annual reports to the Public Service Commission, DPS and the EAG summarizing the results on all programs and from all evaluation studies occurring in the most recent quarter or year. The latest evaluated program savings, realization rates, and net-to-gross ratios will be used in compiling data for these overarching reports. Quarterly reports will be provided to the Commission within 60 days of the end of each calendar quarter. The annual report will substitute for the fourth quarterly report, summarizing program and portfolio progress throughout the calendar year. The annual report will be submitted to the Commission within 90 days of the end of the calendar year.

X. Total Resource Cost Analysis

Once per year, NYSERDA will update benefit/cost ratios (at a minimum, Total Resource Cost test) for each major program and for the entire portfolio of SBC-funded **New York Energy SmartSM** and EEPS programs. The Total Resource Cost (TRC) test divides the present value of the benefits by the present value of Program and Participant Costs. A benefit-cost ratio greater than 1 indicates benefits exceed NYSERDA and participant costs. The Program Administrator Cost (PAC) test divides the present value of the benefits by the present value of the Program Administrator Costs. A benefit-cost ratio greater than 1 indicates benefits exceed NYSERDA costs. For more detailed definition of benefit/cost terms and a description of NYSERDA's current benefit/cost input sources, including avoided energy, capacity and distribution costs, refer to Appendix A of NYSERDA's September 22, 2008 Energy Efficiency Portfolio Standard Program Administrator Proposal.

The latest evaluated program savings, realization rates, and net-to-gross ratios resulting from the evaluation efforts described in this plan will be used in the annual benefit/cost analysis update. NYSERDA will conduct benefit/cost analysis for its programs in a manner consistent with other program administrators, as appropriate. NYSERDA has knowledgeable staff and a tool in place to accomplish benefit/cost analyses for all of its SBC and EEPS programs. NYSERDA is prepared to make adjustments to its current practice should DPS Staff or the EAG decide that alternative methods, tools, or inputs are superior or would foster greater consistency among program administrators.