Public Utility District No. 1 of Snohomish County

Behavior Based Energy Efficiency Pilot Evaluation Final Report

April 8, 2013

Prepared for:
Laura McCrae
Public Utility District No. 1 of Snohomish County
2320 California Street, E2
Everett, WA 98206-1107

The Cadmus Group, Inc.
An Employee-Owned Company • www.cadmusgroup.com

This page left blank.

CADMUS

Prepared by:
Linda Dethman
James Stewart, Ph.D.
Teri Duncan

This page left blank.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Introduction

The Snohomish County Public Utility District (SnoPUD) hired Cadmus to evaluate its Behavior Based Energy Efficiency (BBEE) Pilot program. The BBEE featured a four-week competition period to save energy, followed by a four-month persistence period, among 10 Starbucks stores. SnoPUD's goals for the pilot were to save energy through changing employee behaviors and to develop a measurement and verification (M&V) protocol to assess behavior-based energy-efficiency savings in small commercial buildings. SnoPUD and its partners (the project team) hoped the pilot would inform the design of a larger behavior-based program.

The project team implemented the BBEE Pilot between March 2012 and December 2012. Throughout the entire pilot period, the project team encouraged employee engagement, provided real-time feedback on energy use via in-store dashboards, and provided energy-savings tips and education to each store. The project team's interactions with store employees were greatest during the competition period and decreased considerably during the persistence period.

The project team also developed and implemented their M&V protocol. As part of the M&V protocol, the project team collected whole-building energy use data at the 10 participating stores and for a control group of 58 stores in Puget Sound Energy's neighboring service territory.

SnoPud hired Cadmus as a third-party BBEE evaluator to perform the following tasks:

- Assess the M&V approach the project team used to measure savings.
- Validate the gas and electricity savings during and after the competition period.
- Provide further insights about the ability of the chosen behavioral interventions (that is, competition, feedback, employee engagement, and energy tips) to influence energy savings, and determine whether the participating stores experienced any other benefits.
- Recommend M&V and process improvements based upon the evaluation results.

Research Methods

Cadmus reviewed the project team's M&V data and documentation and met with them to learn more about their evaluation process. Cadmus then assessed the project team's M&V approach and savings estimates, focusing on research design, data collection, model specification and estimation, savings estimation, and reporting.

To develop insights about the pilot's operation, Cadmus reviewed participant feedback and program documents, and interviewed seven members of the project team to learn more about its processes, its influence on energy-efficiency awareness and actions, and the lessons learned. Cadmus used these resources to develop the process conclusions and recommendations in this report.

Measurement and Verification Conclusions

- Overall, the project team used an appropriate M&V approach. The team collected data and conducted a regression analysis of hourly or daily energy use for Starbucks stores, which is the best approach to estimating savings from behavior change.
- Panel regression analysis of electricity use in participant stores indicated that, on average, each participating store saved 528 kWh, or 4.1%, during the competition period. The total electricity savings during the competition period was 5.3 MWh, with a 90% confidence interval of [4.7 MWh, 5.9 MWh].
- Difference-in-differences (D-in-D) panel regression analysis of the energy use of participant and nonparticipant stores indicated that, on average, each participating store saved 274 kWh, or 2.1%, during the competition period. The total electricity savings during the competition period was 2.7 MWh, with a 90% confidence interval of [1.1 MWh, 4.4 MWh].
- The D-in-D regression estimates control for unobservable, naturally occurring efficiency and changes in business activity; these are the best estimates of the pilot savings.
- The average savings during the competition period of the six stores that use natural gas was 7.8 therms; these gas savings were only statistically significant in two stores.
- In the analysis of electricity use during the persistence period, the D-in-D model may be mis-specified, which undermines the validity of the savings estimates. In particular, the model appears to not properly account for electricity demand for space heating and cooling.
- Overall, none of the issues we identified, with the exception of the specification of the intervention vs. control (I vs. C) regression model in the persistence period, was serious enough to question the energy-savings estimates.
- While the M&V approach was appropriate, it could be improved in the following areas:
    - Research design, including the choice of stores and baseline period
    - Model specification, including modeling the effects of weather and occupancy on energy use
    - Accounting for autocorrelation
    - Quantifying the uncertainty of savings estimates
    - Controlling for differences between stores’ average energy use.

The Recommendations section outlines specifics for these improvements.

Process Conclusions

- Overall, the pilot was successful. While energy savings varied among stores and the 5% savings goal was not reached, the pilot illuminated key opportunities and challenges for implementing and measuring the results of workplace behavior energy-efficiency programs.
- The visible and continued support from all levels of leadership is a key influence on the success of behavioral programs. Management attention can help overcome resistance to adoption of new behaviors and help ensure new behaviors persist over time.
- Efficiency efforts may suffer if they interfere with core business goals. At Starbucks, for instance, leaving the floor to attend to the dashboard sometimes conflicted with core duties, such as engaging customers and increasing sales.
- The competition framework received mixed reviews from stores. Some stores were glad to engage in friendly competition while others felt defeated by it from the start.
- The lighter support and lower visibility of pilot activities during the post-competition period made many employees think the initiative had ended. These changes in pilot implementation likely decreased the attention to energy-saving actions.
- Program communications—such as providing energy-efficiency tips, messages from management, or progress reports—are most effective when they use communication channels that are already established in the businesses.
- Stores that saved more energy during the competition period had these characteristics:
    - A strong energy champion leading the pilot
    - A supportive store culture and management (district and/or store manager)
    - Staff members who clearly understood the pilot goals and process
    - Frequent interactions with the dashboard
    - Staff members who were motivated by the friendly competition framework and had strong team spirit
- Stores that saved less energy during the competition generally had characteristics opposite to those that saved more, including:
    - Lack of a strong energy champion
    - Limited or no support from the district or store manager
    - Staff members who did not understand pilot goal or processes
    - Staff members who infrequently checked the dashboard
    - Staff members who were not motivated by the competition framework
    - Employee turnover during competition, particularly in the management positions

Measurement and Verification Recommendations

- ***Employ an experimental design****.* In the future, the project team should consider incorporating more experimental design elements into the pilot. For example, an experimental design with random assignment of stores to the treatment would increase confidence that the estimated energy savings were caused by the program. A randomized control trial was not feasible for the 2012 pilot given its limited scope and the business considerations of participant stores.
- ***Continue to use whole-building energy-use data****.* Whole-building energy-use data captures the entire range of savings impacts. Energy-use data should be collected at least daily to maximize the likelihood of detecting savings.
- ***Verify parallel trends*** *assumption.* The project team should verify the parallel trends assumption upon which the D-in-D regression analysis depends. The Appendix describes how to test this assumption.
- ***Use a longer, more representative baseline period****.* The baseline period should be longer and should be chosen, as much as possible, to represent weather and business conditions in both the competition and persistence periods.
- ***Employ temperature change points****.* The project team should employ temperature change points (cooling degree hours and heating degree hours) with appropriate base temperatures to account for heating and cooling demands on electricity use. Alternatively, the team could employ a generalized additive model that adequately captures temperature effects on electricity use over the whole range of observed temperatures.
- ***Employ appropriate statistical tests for autocorrelation.*** The project team should employ statistical tests, such as the Durbin-Watson test, to account for autocorrelation when it is detected. Feasible Generalized Least Squares would be an appropriate estimation approach.
- ***Use store fixed effects in regressions***. In the pooled and I vs. C analyses, using store fixed effects in regressions is a less restrictive approach to capturing differences between stores' average energy use than the approach the project team did employ.
- ***Adjust aspects of the persistence period electricity savings analysis.*** The project team should verify the equivalence of the intervention and control groups, then re-specify and re-estimate the regression model to account for temperature change points.

Process Recommendations

- ***Secure strong and ongoing management support****.* While the project team secured strong initial support from corporate and mid-management, the pilot would have benefited from a more visible role for mid-managers over time, so that their support would be constantly reaffirmed. For instance, the project could provide mid-managers with success stories and they, in turn, could recognize those successes with their store partners.

- ***Consider the pros and cons of competitions.*** If the project team wants to use competition in the future as a key behavioral driver, more research is needed to determine how well it will engage participants.
- ***Respect and integrate pilot activities with core business goals.*** The project team should ensure that efficiency initiatives are well integrated with core business goals. For instance, since customer engagement is strongly emphasized at Starbucks, store partners might talk with customers about their green efforts or postings in the store might emphasize green activities.
- ***Include energy-efficient behaviors as part of standard operating procedures.*** The project team should investigate if energy efficiency actions can be integrated into operations manuals and other instructions that guide store operations.
- ***Use modes of communication familiar to participants.*** The project team should investigate the normal communications routes that influence employees, and use those routes to deliver new efficiency tips and reminders.
- ***Provide more active support over a longer time period.*** To achieve better persistence of energy saving actions, communications needed to continue over a longer period of time. The project team could support persistence through ongoing communications, feedback, and nudges.

INTRODUCTION

The Snohomish County Public Utility District (SnoPUD) hired Cadmus to evaluate its Behavior Based Energy Efficiency (BBEE) Pilot program, a workplace energy-efficiency behavior-change program.

Program Description

In 2012, SnoPUD partnered with Starbucks, Puget Sound Energy, Lucid, and PECI (the project team) to implement and evaluate the BBEE, which operated in 10 Starbucks stores between July 2012 and December 2012. Through the BBEE, the project team encouraged employees to take energy-saving actions in Starbucks stores, tested the efficacy of inter-store competition, delivered real-time feedback about energy use through Lucent's Building Dashboard® platform, and provided energy-savings education and tips.

The project team identified the 10 stores that participated in the pilot. To be eligible, stores needed to be located in SnoPUD's service territory, and together needed to represent a variety of store types, building vintages, and sizes. The 10 stores included only café and drive-through stores and excluded the smallest and largest stores.

Starbucks informed the selected store managers of the pilot during a conference call on July 23, 2012. The project team presented an overview of the competition and delivered collateral material. The collateral included information about tuning-up the store HVAC system, which could be done at the discretion of the store manager. During the following week, the project team installed an energy-use dashboard in the back room of each store. Also during that week, the project team asked store managers to explain the competition to the store employees, recruit an in-store efficiency champion, introduce the dashboards, and distribute pre-competition surveys.

During the four-week competition period, the project team and the store managers and partners communicated via the energy-use dashboard. The dashboard was designed for two-way communications, providing savings tips and accepting feedback and suggestions. The in-store champions used the dashboard to provide the project team with status reports, including the actions partners took in response to the tips. When the competition period concluded, Starbucks conducted an action review with store managers and partners to gather feedback, insights, barriers, and lessons learned.

The dashboard, however, remained in the stores for another four months. During this persistence period, the project team continued to encourage partners to take efficiency actions through energy-savings tips and performance feedback.

Evaluation Objectives

Working in concert with the project team, Cadmus designed this evaluation to meet the following goals:

- Assess the measurement and verification (M&V) methodology used to measure savings and identify any improvements.
- Validate the electricity and gas savings during and after the competition.

- Provide insights about how much the chosen behavioral interventions (that is, competition, feedback, employee engagement, and energy tips) influenced energy savings and determine other benefits to the participating Starbucks stores. The insights should reflect lessons learned based on feedback by pilot participants and the project team.
- Recommend M&V and process improvements based on the evaluation results.

Research Methods

Cadmus reviewed the project team's M&V data and documentation and met with the project team to learn more about the process. We carefully assessed their approach and savings estimates, focusing on research design, data collection, model specification and estimation, savings estimation, and reporting.

To develop insights about the pilot operation, Cadmus reviewed Starbucks' employee feedback, reviewed pilot materials, and interviewed seven members of the project team: the utility pilot manager, utility pilot co-manager, implementation manager, technology partner, store project lead, store project co-lead, and the regional utility pilot partner. We addressed the following research questions during the interviews:

- How can the reliability and accuracy of behavior-based energy-savings estimates be improved?
- How does the BBEE Pilot influence energy-efficiency awareness, knowledge, attitudes, intentions, and actions among store managers and partners in the participating stores?
- What factors can explain the differences in energy savings between successful, neutral, and unsuccessful stores?
- How can the pilot be improved to increase participant engagement and energy savings?

Organization of the Report

The subsequent chapters of this report cover Cadmus':

- Assessment of the project team's M&V approach
- Validation of M&V savings
- Findings from the key insights analysis
- Conclusions and recommendations for improving future M&V efforts and workplace behavior program delivery

MEASUREMENT AND VERIFICATION METHODOLOGY ASSESSMENT

This section begins with a brief review of Cadmus' M&V approach and assessment activities. We then assess the project team's M&V approach, focusing on the pilot's research design, regression model specification, and estimation of the savings.

Overview of Measurement and Verification Approach

The project team estimated electricity savings in all 10 Starbucks stores (intervention stores) during both the competition and the persistence periods. The team estimated natural gas savings in six of these stores during the competition period. The other four stores did not have significant gas use. The team did not estimate gas savings in the persistence period because data to establish a valid baseline were unavailable.

To estimate the energy savings, the project team collected data on hourly electricity use (all 10 stores) and daily gas use (six stores); hourly weather from the nearest weather station; and store hours, daily sales, and daily transactions.

The project team also collected data on daily electricity use at 58 Starbucks stores (control stores) in Puget Sound Energy's (PSE) service territory to control for naturally occurring energy efficiency and changes in store operations during the competition and persistence periods.

The project team estimated electricity savings in the competition period using three different approaches:

1. Individual store analysis of hourly electricity use
2. Pooled store analysis of hourly electricity use
3. Pooled store analysis—also known as intervention vs. control (I vs. C) analysis—of daily electricity use with a matched comparison group

In all three analyses, the estimate of electricity savings in the competition period was the difference between observed and baseline energy use. The project team employed multiple approaches for estimating savings because it was uncertain which approach would work best.

In the individual store analysis, the project team established baseline energy use in a separate ordinary least squares (OLS) regression of each individual (intervention) store's energy use on outside temperature and a 0-1 indicator for occupancy. For the pooled store analyses, the project team performed a panel regression analysis of the energy use of intervention stores in the baseline and competition periods. The panel regression also included average hourly consumption in the baseline period as an independent variable to control for differences in average consumption between stores. The project team considered but rejected other independent variables, such as store sales and a 0-1 indicator for weekdays.

For the I vs. C analysis, the project team established the baseline energy use with daily energy use data for the 58 control stores. The project team estimated the savings in an OLS difference-in-differences (D-in-D) regression model of daily store electricity use in the baseline and competition periods.

To estimate gas savings in the competition period, the project team used individual store regressions, modeling each store's daily consumption of natural gas as a function of temperature and store occupancy. The baseline period was May 1, 2012, to July 15, 2012.

Table 1 shows the unit of analysis for the competition period electricity and gas analyses.

Table 1. Project Team Analysis of Competition Period Savings

Energy	Individual Store	Pooled Store	Control vs. Intervention
Electricity	Hourly kW	Hourly kW	Daily kWh
Gas	Daily therms	N/A	N/A

For the persistence period, the project team estimated electricity savings using daily electricity use data for intervention and control stores, as well as OLS D-in-D regression.

Measurement and Verification Assessment Activities

The project team provided Cadmus with the following documentation and data:

- Intervention and control store energy use at an hourly or daily frequency, store hours, sales, and weather
- Summaries of the methodologies for individual store, pooled store, and the I vs. C analyses
- Model outputs and savings estimates

Cadmus reviewed the data and documentation, compiled questions about the M&V process and procedures, and made a preliminary assessment. We then met with the project team to obtain answers to our research questions. We then carefully reviewed and analyzed the project team's M&V approach, focusing on research design, data collection, model specification and estimation, savings estimation, and reporting. The following sections describe our findings.

Research Design

The pilot research design concerns the study's research questions and hypotheses, scope (the eligibility, number of subjects, and study duration), how the program impacts will be observed (an experiment, quasi-experiment, natural experiment, simulation, etc.), the sampling approach (simple random, stratified, etc.), and the types of data to be collected. This section reviews and assesses aspects of the BBEE Pilot's research design.

QUASI-EXPERIMENTAL DESIGN

SnoPUD designed and implemented the BBEE Pilot as a quasi-experiment. The project team selected 10 Starbucks stores in SnoPUD's service territory to participate in the pilot on the basis of store type, building vintage, size, and Starbuck's district manager territory. After the pilot period was over, the project team selected a control group of 58 Starbucks stores in PSE's service territory to control for naturally occurring efficiency and changes in store sales. The primary criteria for control group eligibility were the availability of daily energy use data, data quality, heating fuel, and building characteristics. The project team selected a research design with these features to minimize the impact on store operations and to stay within the pilot budget.

While the project team was limited in its choice of a research design, it is still worth considering the impact of the design on the ability to estimate the pilot savings. One issue is unobservable differences in energy use between treatment and control stores. While the project team checked that the control group stores and 10 pilot stores were similar, there may nonetheless have been unobservable differences between them that affected their energy use. These differences can bias the savings estimates and jeopardize the internal validity of the study.

Internal validity refers to whether or not the savings estimates measure the true effects of the program on consumption. For the results to have internal validity, the control stores must represent what the energy use of intervention stores would have been had they not participated. If the control stores are not representative, then the savings estimates will be biased.[1]

The best way to guarantee internal validity is to conduct an experimental design in which subjects are randomly assigned to treatment and control groups.[2] A randomized control design facilitates causal inference about the pilot's impacts and increases confidence that the study's results measured the pilot's true effects.[3]

It was not feasible to implement the pilot as a randomized control trial because of the limited pilot budget and scope and limitations imposed by Starbucks. In the future, however, it may be feasible for SnoPUD to incorporate more experimental design elements into the study. The large number and homogeneity of Starbucks stores may give SnoPUD this opportunity.

WHOLE-BUILDING ENERGY-USE ANALYSIS

The project team collected hourly or daily energy use data for intervention and control Starbucks stores. The project team then matched hourly weather, daily store sales, or store hours to energy use.

---

1 There is also a second issue concerning external validity: whether the results of this study (in which participants were handpicked) can be extrapolated to the rest of Starbucks stores in SnoPUD's service territory. As this project is a pilot, or proof-of-concept, the external validity of the results may be a secondary consideration.

2 See Shadish, Cook, and Campbell, 2002, for a discussion of how research design affects internal validity.

3 While a randomized control trial is the gold standard in evaluation, there are often ways to control for bias in quasi-experimental research designs. Potential approaches include instrumental variables, structural equation modeling, and control function techniques.

The collection of whole-building energy-use data is appropriate for this study. Because the pilot focused on behavior changes that affected multiple electricity and gas end uses, only whole-building data can capture the entire range of such impacts. However, a challenge with whole-building data is that energy savings from behavior is a small percentage of overall energy use (<5%) and is difficult to detect. Therefore, it is best to use high-frequency hourly or daily energy-use data.

Whole-building energy-use data also cannot reveal the sources of specific savings, such as changes in HVAC use, lighting, or plug loads. Depending on future research objectives, end-use metering could help the project team identify other opportunities for savings, the most appropriate energy-savings tips, and the source of specific savings.

EQUIVALENCE OF INTERVENTION AND CONTROL GROUPS

The pilot's quasi-experimental research design is dependent on the equivalence of intervention and control groups. More specifically, the control and intervention stores must have statistically equivalent consumption trends over the estimation period, conditional on BBEE participation and other observed characteristics. In the D-in-D literature, this is known as the parallel trends assumption.[4] It is possible to indirectly test whether or not the parallel trends assumption holds by comparing the consumption of intervention and control stores in the baseline period. The Appendix to this report describes how the project team could implement this test.

As noted above, the project team took steps to minimize the potential for unobservable differences between the groups and to ensure equivalence. Nonetheless, there may have been differences between the intervention and control stores' other energy-efficiency projects. The project team ensured that there were no other energy-efficiency projects in intervention stores in the baseline and competition periods, but this was not possible in control stores. Because the M&V did not account for this possibility, pilot savings estimates may have been biased downward. In the future, an experimental design might limit this potential.

Model Specification and Estimation

This section provides an overview of the technical issues of baseline definition, model specification, and model estimation.

BASELINE PERIOD SELECTION

The baseline period—May 22 or June 6, 2012, to July 15, 2012—was relatively short. In addition, the weather and business conditions during the baseline period may not have been representative of those in the persistence period. These differences have the potential to bias the estimates of the baseline energy use.

4 See Bertrand, Duflo, and Mullainathan, 2004.

The project team intended a longer and more representative baseline period, but was limited by unforeseen scheduling impacts. As part of the M&V, the project team did verify that temperatures in the baseline, competition, and persistence periods were similar. There were, however, several days of extreme temperatures during the persistence period for which there were no analogues in the baseline period.

MODELING TEMPERATURE EFFECTS ON ENERGY USE

To control for electricity use driven by demand for heating and cooling, the project team included either dry-bulb or wet-bulb temperature as an independent variable in the regression equation.[5] The variable entered the model linearly, implying that a degree of change has the same effect at every temperature.

There are, however, two potential issues with modeling temperature's effects on energy use:

1. First, below a certain temperature (for example, 65 °F), a decrease in temperature ceases to have an effect on the demand for cooling. Similarly, above a certain temperature, an increase in temperature ceases to have an effect on the demand for heating. When temperature enters the model as a simple linear function, it cannot capture these change-points. These change points may be important because the BBEE Pilot study area was in a temperate climate and the study period covered warm and cool summertime temperatures and cool fall and wintertime temperatures.
2. Second, the effect of temperature on electricity demand may be non-linear at high temperatures. This has the potential to bias the savings estimates. For instance, if demand for cooling increases at a decreasing rate, then at high temperatures, the estimated baseline and the energy savings will be biased upwards.

A solution would be to including non-linear functions of both heating degree hours (HDH) and cooling degree hours (CDH) in the regression model. The project team could determine the optimal change point for CDH and HDH by examining the R-squared (R^2), model residuals, and coefficient signs for different temperature change points.

AUTOCORRELATION

With high-frequency energy-use data, such as hourly or daily data, the model error term is likely to exhibit autocorrelation. OLS estimation of a regression model with autocorrelated errors results in savings estimates that are unbiased and consistent, but the OLS standard errors will be biased and all inference procedures based on the standard errors may be incorrect.[6] Therefore, it is important to test and account for autocorrelated errors.

---

[5] The project team determined whether to use dry-bulb or wet-bulb temperature by which variable resulted in a higher model R-squared value.

[6] See Greene, 1997 (pp. 586-591).

For the individual store savings analysis, the project team tested for autocorrelation by calculating the sample autocorrelation coefficient and following the American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE) guideline rule that autocorrelation can be safely ignored if the coefficient < 0.5.[7] The project team found evidence of autocorrelation in three stores, which they concluded did not warrant adjusting their approach.

Autocorrelation of 0.5 is significant, and Cadmus was surprised to learn that the threshold in the ASHRAE guidelines is so high. We checked with one of the authors of those guidelines and verified that the 0.5 threshold is a typographical error. The correct threshold should be 0.05.

The most reliable and widely used test of autocorrelation is the Durbin-Watson test, but other tests are available, including the Breusch-Godfrey and Box and Pierce. In addition, it may be possible to detect autocorrelation by plotting the regression residuals against time. Alternating runs of residuals above or below zero suggest that the errors are positively correlated. For panel data such as those used in the pooled or I vs. C analyses, there are specific tests for autocorrelation.

If autocorrelation is found, feasible generalized least squares (FGLS) can be used to estimate the model. This approach results in unbiased and consistent model coefficients and standard errors. FGLS can be implemented in two stages of least squares or in a single step by maximum likelihood. FGLS is available in most statistical software packages. SAS® software includes PROC AUTOREG for individual models and PROC PANEL or PROC TSCSREG for panel models.

DIFFERENCES BETWEEN STORES IN AVERAGE ELECTRICITY USE

In the pooled and I vs. C analyses, it is necessary to control for differences in average electricity use between stores. These differences may arise from the size, type, location, or customer volumes of stores and may be unobservable to the analyst. To the extent these differences are correlated with participation but not controlled for, they have the potential to bias the savings estimates.

The project team accounted for differences in average electricity use by including the average baseline energy use as an independent variable in the regression model. As demonstrated in the Appendix, this is an unnecessarily restrictive approach and can bias the savings estimates. A more appropriate approach for controlling for these differences would be to use store fixed effects. In a fixed-effects model, there is a separate intercept for each store.

In a model with site-fixed effects, it is most appropriate to interact the site-fixed effects with a dummy variable for occupancy to indicate that the store is open, because the energy use of Starbucks stores will vary the most during business hours.

---

[7] See American Society of Heating, Refrigerating and Air Conditioning Engineers, 2002 (p. 108).

MODELING EFFECT OF STORE HOURS ON ENERGY USE

The project team used a 0-1 variable to indicate whether a store was open in an hour and to capture differences in energy use between open and closed hours. This variable takes on the value 1 if the store was open and 0 if closed. In the regression models, the coefficient on the variable had a positive sign, as expected.

Cadmus agrees that parsimony is usually good in modeling but, in this case, there is opportunity to explain more of the variation in energy use over time with some additional modeling. This is because energy use is likely related to hour of the day. For example, we would expect energy use to be higher during peak store hours.

It is possible to allow the effect of store hours to vary by binning the hours of the day and creating separate dummy variables for the bins: 5:00 a.m. to 7:00 a.m., 8:00 a.m. to 10:00 a.m., 10:00 a.m. to 3:00 p.m., etc. This would reduce the amount of unexplained variation in energy use and possibly increase the precision of the savings estimates.

Another enhancement to the model specifications would be to include an interaction variable between store occupancy and the indicator variable for program participation in the competition and persistence periods. In this specification, the coefficient on program participation would measure the savings during hours when the store was closed and the coefficient on the interaction would measure savings during store hours. As the pilot encouraged behaviors that affect energy use during hours when stores were closed, the estimated energy savings in open and closed hour may be very different.

Savings Estimation

In the individual store analysis, the project team estimated electricity and gas savings as the difference between observed and baseline energy use. They estimated baseline energy use as a regression of energy use *in the baseline period* on temperature and an indicator for the store being open. In the pooled and I vs. C analyses, the project team estimated the savings in a regression of energy use *in the baseline and competition periods* on temperature, a store open indicator, and an indicator for the competition period.

Both approaches are valid; however, for consistency, it would be worth considering estimating all of the models using the same approach. One advantage of using a regression of energy use in the baseline and competition periods is that it is easier to estimate standard errors and confidence intervals for the savings estimates.

QUANTIFYING UNCERTAINTY

In the pooled and I vs. C analyses, the project team reported confidence intervals for the estimated savings based on the estimate of the competition or persistence period regression model coefficient and standard error. As noted above, one of Cadmus' concerns is that the standard errors may have been estimated incorrectly because autocorrelation was ignored.

In the individual store analysis, following the International Performance Measurement and Verification Protocol (IPMVP; 2012, Section B-2.2.2, p. 95-96), the project team reported an approximate standard error that underestimates the uncertainty of the savings. In the Appendix, we derived the exact standard error and statistic for testing the statistical significance of the savings.

ESTIMATION OF PERCENT SAVINGS

In the individual store analysis, the project team estimated the percent savings in the competition period as the ratio of the competition kWh savings to energy use in the baseline period. Energy use in the baseline period may not be a valid baseline for energy use in the competition period, however, because of differences between the length of baseline periods, temperature, and business conditions. The most appropriate denominator for estimating percent savings is baseline energy use in the competition period that is estimated with the regression model.

VALIDATION OF MEASUREMENT AND VERIFICATION SAVINGS

This section validates the specific electricity and gas savings estimates for the competition and persistence periods. As noted above, validation of the savings involved reviewing the data, documentation, and outputs of the specific models. Cadmus did not attempt to replicate the results or independently estimate energy savings, since this was not part of the evaluation scope.

Review of Measurement and Verification Savings

COMPETITION PERIOD

Table 2 shows the project team's savings estimates for individual stores during the competition period. Note that the table includes Cadmus' estimate of the percent savings, which we calculated as the ratio of estimated kWh savings to the predicted baseline (this is called Avoided Energy Use by the project team).

Table 2. Competition Period Electricity Savings - Individual Store Analysis

Store Identifier	Estimated Savings (kWh)	% Savings	Savings Statistically Significant at the 5% Level
Store 1	26	0.2%	No
Store 2	1,023	7.1%	Yes
Store 3	1,211	8.9%	Yes
Store 4	-353	-2.6%	No
Store 5	383	3.4%	Yes
Store 6	90	0.6%	No
Store 7	793	4.9%	Yes
Store 8	755	5.8%	Yes
Store 9	282	2.7%	Yes
Store 10	397	4.5%	Yes
Total	**4,606**	**3.5%**	

Sources: Cadmus obtained the savings estimates and approximate standard errors from PECI. Cadmus calculated the percent savings, which, in most instances, is close to PECI's calculated value. Individual store savings do not equal total because of rounding error.

The point estimates of individual store savings in the competition period ranged from -2.6% (Store 4) to 8.9% (Store 3). Nine of 10 savings estimates were positive and seven were statistically significant at the 5% level. The percent savings across all 10 stores was 4.6 MWh, or 3.5% of energy use in the competition period.

The pooled store analysis of electricity savings in the competition period yielded a similar savings estimate. Table 3 shows those savings estimates with 90% confidence intervals from the pooled and I vs. C electricity analyses.

Table 3. Competition Period Electricity Savings – Pooled and I vs. C Analyses

Model	Per-Store Competition Savings (kWh)	90% CI* Lower Bound	90% CI* Upper Bound	% Savings	Pilot Savings (MWh)	90% CI* Lower Bound	90% CI* Upper Bound
Pooled	528.6	469.2	588.1	4.0%	5.3	4.7	5.9
I vs. C	273.6	107.5	439.7	2.1%	2.7	1.1	4.4

Notes: All of the numbers in this table are from Cadmus calculations based on project team regression outputs.
* CI indicates confidence interval.

The average store electricity savings in the competition period was approximately 530 kWh, or 4%, using the estimated energy use from the individual store analysis as a baseline. The pilot savings in the competition period was approximately 5.3 MWh. The 90% confidence interval of [4.7 MWh, 5.9 MWh] does not include zero, indicating that the savings are statistically significant at the 10% level.

Neither the individual store nor the pooled store analyses control for naturally occurring efficiency or changes in business activity between the baseline and competition periods. After controlling for these factors in the I vs. C analysis, the estimated net savings were still large and statistically significant, but were substantially reduced.

The estimate of per-store net savings is 274 kWh, or 2.1% of baseline consumption. The pilot net savings are 2.7 MWh, with a 90% confidence interval of [1.1 MWh, 4.4 MWh]. The individual store point estimate of pilot savings in the competition period is well outside of this confidence interval, suggesting that the individual store analysis did not account for some changes in energy use that were correlated with program activity.

Table 4 shows the project team's estimates of natural gas savings for six Starbucks stores during the competition period. The project team estimated the savings with regressions of daily gas use on temperature and store sales.

Table 4. Competition Period Natural Gas Savings - Individual Store Analysis

Store	Estimated Savings (therms)	% Savings	Savings Statistically Significant at the 5% Level
Store 3	29.8	39.8%	Yes
Store 4	-9.0	-19.3%	No
Store 5	0.2	0.7%	No
Store 6	16.0	19.4%	Yes
Store 9	11.4	14.9%	No
Store 10	-1.7	-4.3%	No
Total	**46.7**	**13.1%**	

Sources: Cadmus obtained the savings estimates and approximate standard errors from PECI. Cadmus calculated the percent savings.

The individual store gas savings estimates ranged from -19.3% (Store 4) to 39.8% (Store 3). Four of six savings estimates were positive; two of those four were larger than twice the standard error of the regression model. The gas savings are less precisely estimated than the electricity savings, in part because the gas use data are daily instead of hourly. The savings for all six stores was 47 therms, or 13.1% of baseline consumption.

The project team did not estimate pooled or I vs. C analyses of natural gas savings because of the small number of stores.

PERSISTENCE PERIOD

The project team employed a D-in-D analysis to estimate the kWh savings in the persistence period. Table 5 reports an estimate of the per-store and pilot savings in the persistence period.

Table 5. Persistence Electricity Savings – I vs. C Analysis

	Per Store Competition Savings (kWh)	90% CI Lower Bound	90% CI Upper Bound	Pilot Savings (MWh)	90% CI Lower Bound	90% CI Upper Bound
I vs. C	-1,439.7	-1,874.3	-1,005.1	-14.4	-18.7	-10.1

The project team estimated the per-store savings in the persistence period to be approximately -1,440 kWh, and the pilot savings to be 10 times greater.

Cadmus agrees with the project team that the savings are not valid for several reasons:

1. First, the baseline period of May to July does not represent weather or business conditions in the persistence period of September to December. For example, in the baseline period, average temperatures were warmer, and Starbucks likely served fewer hot drinks than in the persistence period.

2. Second, the regression model is not specified correctly. It includes dry-bulb temperature, but does not account for the heating and cooling change points described in the methodology assessment. As specified, the model would predict that electricity use in an electrically heated store would decrease with colder temperatures.[8]
3. Third, the comparison of 58 control stores may not be particularly well matched to the intervention stores during the heating season if there were differences in the proportion of stores using electricity for space heating. The project team should verify a balance in heating fuel types.

Savings Validation

The Measurement and Verification Methodology Assessment section mentioned opportunities for improving this pilot's research design, model specification and estimation, and energy-savings estimation. Specific recommendations for these improvements are described in the Conclusions and Recommendations section below.

Overall, Cadmus determined that the pilot M&V approach was appropriate, and none of the issues we identified, with the exception of the specification of the I vs. C regression model in the persistence period, was serious enough question the energy-savings estimates.

Cadmus determined that the I vs. C electricity savings estimate of 2.1% for the competition period has the most validity, because the savings account for naturally occurring efficiency and changes in business activity in intervention stores. The pilot realized 42% of the goal of 5% savings during the competition period.

The M&V analysis of electricity savings in the persistence period is not valid, however. Cadmus has concerns about the equivalence of the control and intervention groups and about the specification of the I vs. C regression model.

---

8 Only one of the 10 stores was electrically heated, and the project team acknowledged this limitation of their analysis.

KEY INSIGHTS ANALYSIS

Overview

This section documents Cadmus' assessment of the key factors influencing store engagement and energy savings, based on our review of employee feedback, program materials, and most prominently, the in-depth interviews with the project team. The interview guide is included in the Appendix.

The overall research questions for the key insights analysis were:

- What were the motivations for participating? What were the expectations of participants?
- What energy and non-energy benefits and challenges did the pilot program produce within stores?
- What are the factors that influenced savings and persistence of savings?

Delivery of the Pilot Program

The project team members agreed that the written program materials accurately describe what actually happened, and that no significant differences exist between what was written and actual pilot operation. One team member said: "*Okay, we did it the way we said we were going to, these are results, and here's what we will do next time.*"

GOALS OF THE PILOT

When asked to tell us about the goals of the BBEE, one project team member said: "*the first three goals [listed below] are just as important as the energy savings*" and the energy savings, while important, were: "*not the only objective*" of this program. Overall, the team reported that the pilot had these goals:

- ***Engagement.*** Engage with: "*small retail/commercial customers and their employees*."
- ***Scalable.*** "*Develop a scalable behavioral energy-efficiency program model; whatever we did with a company like Starbucks could be scaled regionally or scaled across disparate businesses.*"
- ***Develop an M&V Approach.*** Develop an M&V approach to apply to small commercial facilities ("*no protocol exists for small commercial facilities*") that could be evaluated and: "*standardize process so that other utilities can use it for focusing dollars towards behavioral energy efficiency.*"
- ***Energy Savings.*** Demonstrate 5% energy savings during the 30-day competition.
- ***Insight.*** Gain a better understanding of how to implement behavioral energy-efficiency programs in a small commercial/retail setting ("*[we have made] many touches to small commercial [customers]; none have been very effective*").

Program Process and Flow

Members of the project team were consistent across all of their interview responses, with some providing insights about operations where others were less involved. Many steps of the operations happened concurrently, while the project team had to add some steps quickly to ensure that the pilot adhered as closely as possible to its plan. The key process steps are listed in Table 6 below.

Table 6. Key Process Steps in Pilot Program Delivery/Operation

Before Event Kickoff	Event (Competition) Launch
• Develop an operational program plan • Develop an M&V plan and project schedule • Select test and control group stores • Research the Lucent dashboard functions and interactions • Develop protocol for data collection • Set up the information technology (IT) interface • Coordinate meter upgrades (unforeseen step) • Install tablets/dashboard communication in stores (the unforeseen needs related to this were: addressing IT issues with Wi-Fi and the need to mount tablets on the wall to function in a back-of-store environment) • Develop M&V methods (none existed prior to this pilot for small commercial BBEE programs) • Conduct field visits to nonparticipating Starbucks stores to learn the operational environment • Develop messaging and materials for dashboard and competition	• Host kickoff meeting with managers one week before competition • Launch competition and collect data • Wrap up the competition • Continue to collect data, disable dashboard internet (e-mail) function, continue dashboard function only • Pick up dashboard and collect data • Conduct M&V activities: data coordination and cleaning, analysis, etc. (data accessibility and availability was an issue for some meter data.)

Influencing Factors for Success

We asked the project team what indicators should be used to measure the pilot's success. This section summarizes their perspectives on the components and activities that contributed to the successes and challenges of the pilot.

OVERALL PILOT SUCCESSES

Overall, the project team agreed that the pilot was a success and that it produced valuable insights for future projects, although they would have liked to see stronger persistence data associated with behavior changes ("*…it's so new it's bound to have issues and [we] need to do this a couple of times to learn more to do it better. Lots of potential*"). They thought the energy-saving actions and information provided to Starbucks employees was relevant, simple, and successful. The project team said that they gained invaluable insight on how to effectively coordinate using the Lucent dashboard, installing and using advanced metering infrastructure (AMI) metering data, and balancing program design and messaging with customer priorities and communication channels.

Starbucks' corporate attitudes were consistent with the pilot's goals, and the sound working relationships the project team formed during previous energy-efficiency projects made it easy for

them to work together. Some of the project team members developed the pilot idea and approached SnoPUD, after identifying it as a "*progressive utility*" and a good fit. Starbucks already had long-term sustainability goals in place. A regional utility partner said they: "*see behavior-based energy efficiency as a primary source of savings going forward, [and believes that] this pilot has been successful in identifying elements [of tracking and monitoring] needed…[for] capturing savings.*"

The project team stated that their communicative and collaborative style facilitated a solutions-oriented approach to navigating many logistical barriers. For instance, when the team encountered barriers to interfacing the meter data with the dashboard technology, the regional utility partner expedited the installation of upgraded meters. Several project team and partners echoed the sentiment of one partner who said: "*I've done lots of pilots, there is a real strength in this one; everyone was into it for common good.*"

The project team said one reason Starbucks was an ideal partner for a workplace behavior energy-efficiency pilot was because its standard operating procedures reduced the need for individual store adjustments. For instance, each Starbucks location uses standardized procedures for store operations—such as for daily opening tasks, cleaning, reporting, and daily closing. However, the project team did state that pilot implementers needed to allow enough time to navigate and understand Starbucks' protocols.

The pilot was successful in developing simple, relevant behavioral energy-efficiency actions and tips for employees in small commercial/retail environments. The store managers and the project team were happy with the simplicity and relevance of the energy-saving tips. Store managers stated that the energy-saving actions should be included with existing operation checklists and procedures to make it less of an optional or additional thing for their employees' to do.

The project team effectively developed the first behavior-based M&V protocol for small commercial customers. The project team said that while they spent considerable time trying several M&V models and protocols, it was worth it because the protocol will: "*pave the way*" for other utilities that wish to provide similar programs ("*the M&V was more difficult than anticipated. I feel good about the M&V that was done; it's going to deliver value to region*").

KEY FACTORS FOR SUCCESSFUL STORES

The project team identified key factors for successful stores; that is, those that saved the most energy. The most successful stores had the right mix of individuals, motivation, and pilot engagement, including these key factors:

- ***An active energy champion and an engaged, enthusiastic store manager.*** Stores with an engaged store manager and/or a designated energy champion demonstrated higher levels of savings ("*having a leader in the store [who was] enthusiastic helped; best case—if the leader was a store manager*"). Store managers had the option to either lead the effort or designate a champion to: "*communicate tactics and tips to [the rest of the] store*" employees. Stores that designated a store champion (and were highly engaged with the dashboard, as discussed below) were: "*far and away more engaged than other stores*."
- ***Strong employee interaction with the dashboard.*** Since the dashboard was the: "*main source of information for comparing performance with other stores*" and was the main mode of communication overall, stores that had regular and active interaction with the dashboard were most successful. The project team added that: "*if the store had people that were more interested in engaging through social media and electronic communication, [then they] were more successful.*"
- ***Employees being motivated to engage in friendly competition and having a strong team spirit.*** One of the project team members said: "*there is a real team environment in the [Starbucks'] stores*" making them: "*well suited for a competition*." Another said: "*[This type of program] needs leaders to really take it on; some stores really had that [while] others didn't.*" Successful stores had employees who were willing to not only take the suggested actions, but who came up with additional suggestions for long- and short-term ideas to further the stores energy savings. One project team member said the winning store was: "*very actively looking to problem solve and work out [their] own ways to save energy.*"

KEY CHALLENGES TO SUCCESS

The key factors that likely affected store success are:

- ***Lack of an energy champion or enthusiastic and engaged store manager.*** Stores that experienced limited success with the program did not have a champion or supportive store manager to motivate employees ("*if there wasn't a champion, or [employees weren't] directed by the store manager, then [that store was] not as successful*").
- ***Limited interaction with the dashboard.*** Less successful stores had limited interaction with the dashboard and the energy-saving tips provided, making it hard to attribute any change to the pilot. One project team member described the range in energy savings as being between 0.5% and 10% savings, and that: "*the low percentages seemed [more] consistent with weather changes, as opposed to the more active, engaged stores with 5-10% savings*."

- ***Lack of interest in competing.*** Highly successful stores may have discouraged participation at other stores. The employees in an unsuccessful store may have felt they "*didn't have a chance*" in the competition so became focused on core business goals instead.
- ***Management directives to focus on core business priorities.*** Strong direction from corporate management to focus on core business objectives caused some stores to deprioritize the pilot.
- ***Employee and management turnover.*** Turnover in both district managers and some store managers inevitably disrupted the continuity of the enthusiasm and focus on the competition ("*the shift [in managers] changed the focus*").

Lessons Learned

The key lessons from this pilot focused on better matching the pilot design to the customers' existing communication channels and operations environment and allowing enough time for the project team to merge data from disparate metering technologies and the dashboard.

The project team provided the following lessons learned from the BBEE:

- ***The customers' appropriate communication channels need to be learned, mapped, and established for pilot use.*** The project team and Starbucks worked together to determine appropriate communication channels for the project without disrupting store operations; however, the project team said they should have allowed more time to achieve buy-in throughout the entire process, not just in the beginning stages. Having an engaged key point of liaison at Starbucks was critical to the pilot's success. The management at Starbucks also provided information on what the company sustainability goals, which the liaison helped connect with the pilot effort.
- ***Employees needed to know the parameters of the competition.*** Despite the team's efforts at messaging, employees believed that the pilot effort did not extend beyond the 30-day competition ("*employees really felt the initiative was just for that one month*"). Many project team members stated: "*we needed more ongoing communication*" beyond the dashboard. One member said they did not foresee that: "*super engaged [employees]*" might take actions the company did not want, such as: "*shutting down the espresso machines*" in order to boost their place in the competition.
- ***Energy-saving behaviors can be institutionalized in regular duty checklists.*** The team learned that Starbucks likely already had intersections where operations and employee behavior met, as described by one team member who stated: "w*e currently have a tool, the duty roster, that contains [daily and monthly] cleaning schedules. We could put a few tips in there. Partners initial the roster when they complete the task. They are familiar with it. Instead of introducing a brand new tool, use something like a duty roster for savings tips.*"
- ***The back-of store environment for the dashboard did not fit all stores.*** The project team recognized that the most successful stores used the dashboard; however, for some stores, it was not a familiar mode of communication and was not used. One team member stated: "*the*

dashboard became a new thing [to look] at one time and forget. We didn't have ongoing dialog. We need to figure out a way to have ongoing dialog in ways [employees] are used to." Another project team member said: "*I wouldn't put an extra piece of hardware in the store. Also, based on cost, based on payback in terms of savings, it is not worth it.*" In addition, several team members noted it was important to have multiple modes of communication.

- ***Tips had to be simple and relevant to be effective.*** These messages' simplicity fit well into normal business operations, manuals, and operational reminders. One member stated: "*we think the tips we developed were valid and easy and should be the way we are doing work here rather than a temporary initiative. We're looking at developing it as standard practice [in] how we do things. [We will] Also incorporate the point of use tips, like how to use the dishwasher with the least environmental impact.*" The project team also learned where to place the energy-saving tips for employees: "*we learned some things about the technology; mounting [the tablet] on the wall is not as impactful as it being where they already have access to information, like their cell phones, or their weekly/daily logbook, bulletin boards, manager e-mails, or employee portal.*"
- ***Coordinating data collection was complex.*** Coordinating data collection from three different utilities with different meter systems was much more time consuming and complex than anticipated: "*[we] learned that there is a lot to consider*". Several project team members stated they were anticipating less difficulty extracting and sharing data from the new meters and meters from other utilities. The project team also would have liked an extended baseline collection period and better baseline data: "*the biggest lesson: plan for extended baseline data collection period, or ensure we have ways to get improved baseline data collection.*"
- ***M&V was more complicated and took longer than expected.*** The project team said they spent a lot of time trying several M&V models and the dashboard did not offer M&V options. One team member added: "*we needed to get data from the dashboard then put it into SAS, and the same for the utility data—would be nice to get data all in one place.*"

PERSISTENCE OF BEHAVIOR CHANGE FROM THE PILOT

Most project team members agree that Starbucks' energy-savings behaviors are unlikely to persist without further engagement. They said if the new energy-saving tips become part of standard operating procedures, and appear where employees are accustomed to accessing direction or information, energy-savings behaviors are more likely to persist.

However, the team said the pilot tested the impact of a single mode of communication—the dashboard—on behavior within a 30-day energy-savings competition and a four month persistence period. Messaging continued after the competition ended, but other interaction ceased. In addition, the pilot did not clearly set any longer-term expectations.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

OVERALL MEASUREMENT AND VERIFICAITON APPROACH

Based on review and assessment of the project team's M&V methodology and the energy-savings estimates, Cadmus concludes the following:

- Overall, the project team used an appropriate M&V approach. The collection and regression analysis of hourly or daily energy-use data for Starbucks stores is the best approach to estimating savings from behavior change.
- None of the issues we identified, with the exception of the specification of the I vs. C regression model in the persistence period, was serious enough to question the energy-savings estimates.
- While the M&V approach was appropriate, it could be improved in the following areas:
    - Research design, including the choice of stores and baseline period
    - Model specification, including modeling the effects of weather and occupancy on energy use
    - Accounting for autocorrelation
    - Quantifying the uncertainty of savings estimates
    - Controlling for differences between stores' average energy use

The Recommendations section outlines specifics for these improvements.

ELECTRICITY AND GAS SAVINGS ESTIMATES

Cadmus based the following findings on our review and validation of electricity and gas savings estimates:

- Panel regression analysis of electricity use in participant stores indicated that, on average, each participating store saved 528 kWh, or 4.1%, during the competition period. The total electricity savings during the competition period was 5.3 MWh, with a 90% confidence interval of [4.7 MWh, 5.9 MWh].
- D-in-D panel regression analysis of the electricity use in participant and nonparticipant stores indicated that, on average, each participating store saved 274 kWh, or 2.1%, during the competition period. The total electricity savings during the competition period was 2.7 MWh, with a 90% confidence interval of [1.1 MWh, 4.4 MWh].
- The D-in-D regression estimates control for unobservable, naturally occurring efficiency and changes in business activity; these are the best estimates of the pilot savings.

- The average savings during the competition period of the six stores that use natural gas was 7.8 therms; however, these gas savings were only statistically significant in two stores.
- In the I vs. C analysis of electricity use during the persistence period, the D-in-D model is possibly mis-specified, and there may be significant differences between intervention and control stores that undermines the validity of the savings estimates. In particular, the model appears to not properly account for electricity demand for space heating and cooling.

KEY INSIGHTS ANALYSIS

Based upon our BBEE insights analysis, Cadmus reached the following conclusions:

- Overall, the pilot was successful. While energy savings varied among stores and the overall 5% savings goal was not reached, the pilot illuminated key opportunities and challenges for implementing and measuring the results of workplace behavior energy-efficiency programs.
- The visible and continued support from all levels of leadership is a key influence on the success of behavioral programs. Management attention can help overcome resistance to adoption of new behaviors and help ensure new behaviors persist over time.
- Efficiency efforts may suffer if they interfere with core business goals. At Starbucks, for instance, leaving the floor to attend to the dashboard sometimes conflicted with core duties, such as engaging customers and increasing sales. Similarly, energy-efficient behaviors that are integrated as much as possible in required daily routines (for instance, part the operating manual or daily checklists) are more likely to persist.
- The competition framework received mixed reviews from stores. Some stores were glad to engage in friendly competition while others felt defeated by it from the start.
- The lighter support and lower visibility during the post-competition period made many employees think the initiative had ended. These changes in pilot operation likely decreased the attention to energy-saving actions.
- Program communications—such as providing energy-efficiency tips, messages from management, or progress reports—are most effective when they use communication channels that are already established in the business.
- Stores that saved more energy during the competition period had these characteristics:
    - A strong energy champion leading the pilot
    - A supportive store culture and management (district and/or store manager)
    - Staff members who clearly understood the pilot goals and process
    - Frequent interactions with the dashboard
    - Staff members who were motivated by the friendly competition framework and had a strong team spirit
- Stores that saved less energy during the competition generally had characteristics opposite to those that saved more, including:

- Lack of a strong energy champion
- Limited or no support from the district or store manager
- Staff members who did not understand pilot goal or processes
- Staff members who infrequently checked the dashboard
- Staff members who were not motivated by the competition framework
- Employee turnover during competition, particularly in the management positions

Recommendations

MEASUREMENT AND VERIFICATION RECOMMENDATIONS

- ***Employ an experimental design***. In the future, the project team should consider incorporating more experimental design elements into the pilot. For example, an experimental design with random assignment of stores to the treatment would increase confidence that the estimated energy savings were caused by the program. A randomized control trial was not feasible for the 2012 pilot given its limited scope and the business considerations of participant stores.
- ***Continue to use whole-building energy-use data***. Whole-building energy-use data captures the entire range of savings impacts. Energy-use data should be collected at least daily to maximize the likelihood of detecting savings.
- ***Verify parallel trends assumption.*** The project team should verify the parallel trends assumption upon which the D-in-D regression analysis depends. The Appendix describes how to test this assumption.
- ***Use a longer, more representative baseline period***. The baseline period should be longer and should be chosen, as much as possible, to represent weather and business conditions in both the competition and persistence periods.
- ***Employ temperature change points***. The project team should employ temperature change points (CDHs and HDHs) with appropriate base temperatures to account for heating and cooling demands on electricity use. Alternatively, the team could employ a generalized additive model that adequately captures temperature effects on electricity use over the whole range of observed temperatures.
- ***Employ appropriate statistical tests for autocorrelation.*** The project team should employ statistical tests, such as the Durbin-Watson test, to account for autocorrelation when it is detected. Feasible Generalized Least Squares would be an appropriate estimation approach.
- ***Use store fixed effects in regressions***. In the pooled and I vs. C analyses, using store fixed effects in regressions is a less restrictive approach to capturing differences between stores' average energy use than the approach the project team did employ.
- ***Use exact standard errors in the individual store analysis.*** The project team employed approximate standard errors that underestimated uncertainty in testing the statistical

significance of savings. This could lead to incorrect inferences about the significance of store savings.

- ***Adjust aspects of the persistence period electricity savings analysis.*** The project team should verify the equivalence of the intervention and control groups, then re-specify and re-estimate the regression model to account for temperature change points.

PROCESS RECOMMENDATIONS

- ***Secure strong and ongoing management support***. While the project team secured strong initial support from corporate and mid-management, the pilot would have benefited from a more visible role for mid-managers over time, so that their support would be constantly reaffirmed. For instance, the project could provide mid-managers with success stories and they, in turn, could recognize those successes with their store partners.
- ***Consider the pros and cons of competitions.*** If the project team wants to use competition in the future as a key behavioral driver, more research is needed to determine how well it will engage participants.
- ***Respect and integrate pilot activities with core business goals.*** The project team should ensure that efficiency initiatives are well integrated with core business goals. For instance, since customer engagement is strongly emphasized at Starbucks, store partners might talk with customers about their green efforts or postings in the store might emphasize green activities.
- ***Include energy-efficient behaviors as part of standard operating procedures.*** The project team should investigate if energy efficiency actions can be integrated into operations manuals and other instructions that guide store operations.
- ***Use modes of communication familiar to participants.*** The project team should investigate the normal communications routes that influence employees, and use those routes to deliver new efficiency tips and reminders.
- ***Provide more active support over a longer time period.*** To achieve better persistence of energy- saving actions, communications needed to continue over a longer period of time. The project team could support persistence through ongoing communications, feedback, and nudges.

REFERENCES

American Society of Heating, Refrigerating and Air Conditioning Engineers. *ASHRAE Guideline: Measurement of Energy and Demand Savings*. 2002.

Bertrand, Marianne, E. Duflo, and S. Mullainathan. “How Much Should We Trust Difference-in-Differences Estimates.” Quarterly Journal of Economics (2004): 119 (1). pp. 249-275.

Greene, William. *Econometric Analysis (6th edition).* 1997. New York: Prentice-Hall.

Efficiency Valuation Organization. *International Performance Measurement and Verification Protocol: Concepts and Options for Determining Water and Energy Savings.* 2012.

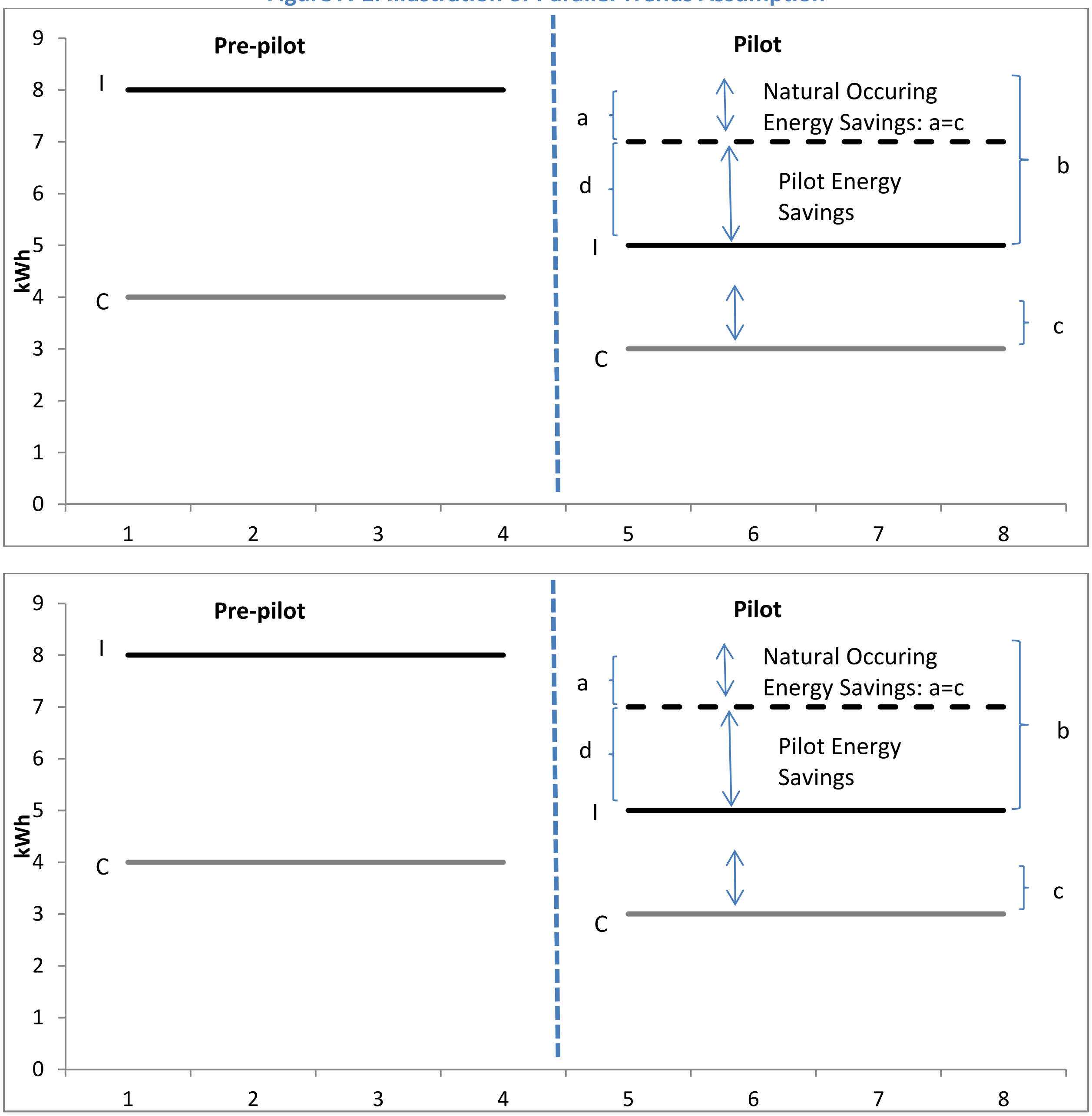
Shadish, William R., T.D. Cook, and D.T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth. 2002.

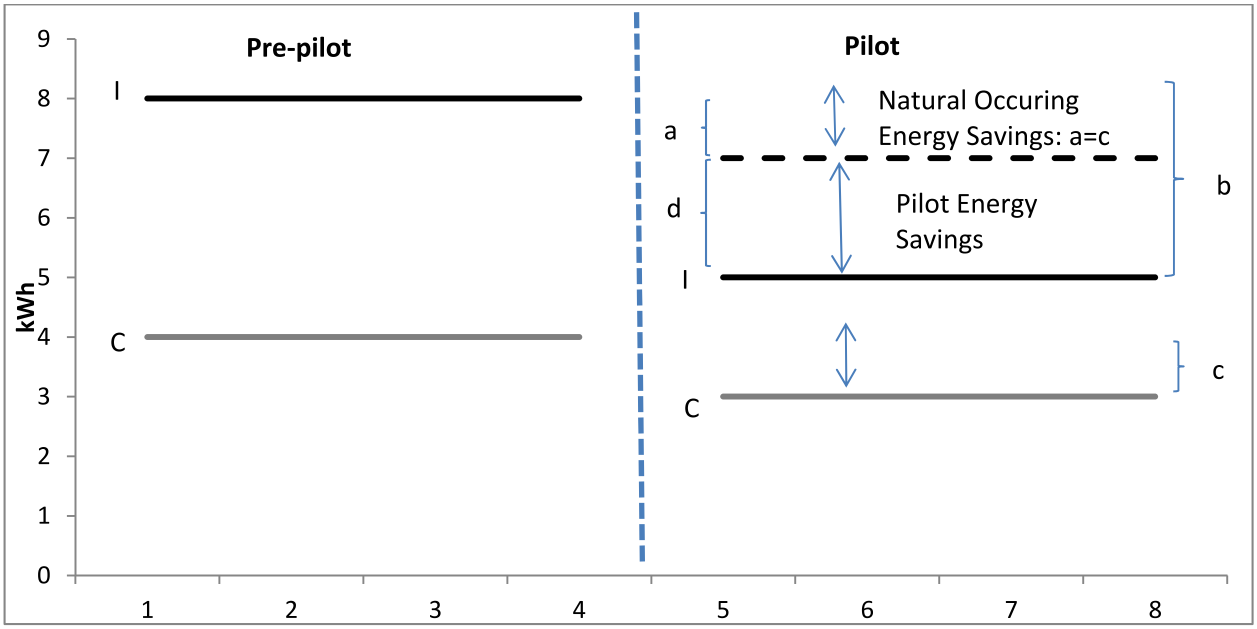
APPENDIX

Parallel Trends Assumption

Identifying the pilot savings using D-in-D regression requires that intervention and control stores followed parallel trends in electricity consumption over the estimation period, conditional on their participation and other observed characteristics. Figure A-1 illustrates this assumption.

Figure A-1. Illustration of Parallel Trends Assumption





The figure shows the conditional mean consumptions of a hypothetical intervention site and a control site before and after the start of the pilot. The solid black line and the letter '*I*' indicate the consumption of the intervention site. The solid gray line and the letter '*C*' indicate the consumption of the control site. The vertical dashed line denotes the start of the pilot.

Using the figure as an example, assume there are two ways a site can reduce its energy use in the pilot period: pilot participation or naturally occurring (independent) behavior change. First, consider the intervention site. Before the pilot, energy use is 8 kWh/period. (In this example, all references to kWh refer to energy use per period.) After the pilot start, the intervention site's energy use decreases to 5 kWh. Hence, the total energy use decreases by 3 kWh (shown as the difference '*b*').

However, some of this energy savings is due to naturally occurring behavior change (that is not observable and directly measurable). Suppose that if the pilot had not occurred, the site's consumption would have decreased by 1 kWh to 7 kWh. This implies that the initiative energy savings equal 2 kWh (shown by the difference '*d*').

In a simple pre-post analysis of consumption, it would not be possible to estimate the pilot energy savings, because savings from the pilot and naturally occurring behavior change cannot be identified separately. Energy savings would incorrectly be estimated as 3 kWh.

Now suppose there is a site that did not participate in the pilot. This control site had lower energy use before the pilot than the participant site. Also, suppose that *whatever naturally occurring behavior changes occurred in the intervention site also occurred in the control site*. This is the parallel trends assumption. Then it is possible to use the change in energy use at the control site to identify the pilot savings, by subtracting the control site's energy savings (shown by '*c*') from the intervention site's change in consumption (the D-in-D).

The ability to estimate the pilot net savings depends on the parallel trends assumption. If naturally occurring adoption in the control site was more than that in the intervention site and resulted in more savings (for example, if it was 3 kWh instead of 2 kWh), the assumption of the D-in-D approach would be violated, and the estimate of pilot savings would be biased.

To check the parallel trends assumption, trends in intervention and control consumption before the start of the pilot should be examined. If the two groups exhibited similar trends in consumption before the pilot, it is likely they would have followed similar trends after the pilot. Specifically, this could be accomplished by estimating a store and time fixed-effects regression model of daily consumption (see Equation A-1).

Equation A-1. Parallel Trends Regression Equation

$$\ln\left(\mathrm{DC}_{i,t}\right) = \alpha_i + \sum_{d=2}^{D} \mu_d \cdot \mathrm{Day}_d(t) + \boldsymbol{X}_{i,t}^{\mathrm{T}}\boldsymbol{\beta} + \theta \cdot \tau(t) + \theta_p \cdot \tau(t) \cdot \mathrm{Int}_i + \varepsilon_{i,t}$$

For store '*i*' and period '*t*' (where each period equals one day), the equation definitions are as follows:

α_i	=	Intercept term specific to store '*i*.' This corresponds to the store's average hourly load.
$\mathrm{Day}_d(t)$	=	Indicator variable indicating whether period '*t*' is in day '*d*.'
$\boldsymbol{X}_{i,t}$	=	Vector of additional explanatory variables. The additional explanatory variables used in this analysis are store hours and CDHs and HDHs.
$\tau(t)$	=	Time trend variable that increases by one with each period. The time trend captures the amount by which daily consumption increases or decreases in each month. For this analysis, we defined $\tau(t)$ as the number of days from the start of the baseline period.
Int_i	=	Indicator variable for participation in the BBEE Pilot. This variable equals one if the site was an intervention store (and equals zero otherwise).

This panel equation should be estimated with energy-use data for intervention and control stores before the pilot started. The coefficient 'θ_p' on the interaction between 'Int_i' and the time variable 'τ' represents the average difference between intervention and control site consumption trends after controlling for differences in time-invariant consumption, day-specific consumption effects, and the explanatory variables included in '$\boldsymbol{X}$.'

Using Site-Fixed Effects to Account for Differences in Average Energy Use

To see the difference between site-fixed effects and a variable for average energy use in the baseline period, suppose there are two stores, A and B, and two periods, 1 and 2. A has average energy use of α_A per period and B has average energy use of α_B per period. Both α_A and α_B are unknown and must be estimated. Also, in period 2, B participates in an energy-efficiency program, which has an unknown effect $\theta<0$ on energy use, kWh. Let D=1 if the store participates and 0, otherwise. For simplicity, assume all are other factors affecting energy use, ε, are random and unobservable with expectation of zero. Thus, letting s index the store and t the period, the true model is:

$$kWh_{st} = \alpha_s + \theta D_{st} + \varepsilon_{st}$$

where α_s is the unobservable average consumption for store s. For example, the expected consumption of store A in period 2 is $E[kWh_{A2}|D_{A2}=0]= \alpha_A$ and the expected consumption of store B in period 2 is $E[kWh_{B2}|D_{B2}=1]= \alpha_b + \theta$.

Now, consider the project team's model:

$$kWh_{st} = \gamma kWh_{s1} + \theta D_{st} + \varepsilon_{st}.$$[9]

The expected consumption of store A in period 2 is $E[kWh_{A1} \mid kWh_{A1}, D_{st}] = \gamma kWh_{A1}$, where γ is the average effect of period 1 consumption on period 2 consumption across stores. This model will result in an unbiased estimate of consumption only if

$$\gamma kWh_{A1} = \alpha_a, \text{ or, if } \gamma = \alpha_a / kWh_{A1}.$$

For B, the necessary condition is $\gamma = \alpha_b/kWh_{B1}$.

Thus, it is necessary that $\alpha_a/ kWh_{A1}= \alpha_b/kWh_{B1}$ to obtain unbiased savings estimates of period 2 consumption for A and B. This condition says that the ratios of unknown average consumption to observed period 1 consumption for A and B must be equal.

With a store fixed-effects model, this condition need not hold to achieve unbiased estimates.

Testing the Statistical Significance of Individual Store Savings

In the individual store analysis, energy savings in the competition and persistence periods is defined as the cumulative difference between baseline energy use and observed energy use, where baseline energy use is predicted using the coefficients of a regression of energy use in the baseline period. Baseline energy use is estimated under the null hypothesis that the intervention did not influence energy use, that is, savings were zero, or equivalently, that there was no intervention. If the difference

[9] We suppress the intercept to approximate the project team model: there is one intercept for 10 participating stores.

between observed and baseline energy use is negative and large, this would be evidence of energy savings.

Suppose the true but unknown regression model of energy use in the baseline period is:

$$kWh_t = \mathbf{X}_t'\beta + \varepsilon_t \qquad (1)$$

where X_t is a 1 x k vector of k-1 explanatory variables and a constant, β is a vector of coefficients for the variables in $\mathbf{X}_t$, and ε_t is an error term that is normally, independently, and identically distributed with mean zero and variance σ^2. Using data from t=1, 2, …, T periods during the baseline period (before the pilot), the project team estimated the model by OLS and obtained estimates of the vector β and error variance σ^2, denoted b and s^2.[10] The estimate of the variance-covariance matrix for b is $s^2(\mathbf{X}'\mathbf{X})^{-1}$, where X is a T x k matrix.

Next, the project team used the model $kWh_t\text{^} = \mathbf{X}_t'\mathbf{b}$ to predict energy use in the competition period under the hypothesis that the intervention had no effect. For each of the t=1, 2, … , T^c periods during the competition, we observe $\mathbf{X}_t$, which we denote $\mathbf{X}_t^C$ to distinguish it from the baseline period values.

The energy savings in period t, an hour or a day, of the competition are:

$$ES_t^C = kWh_t^C - kWh_t^C\text{^} = \varepsilon_t^c - \mathbf{X}_t^{C\prime}(\mathbf{b} - \beta)$$

where ES_t^C is the estimated energy savings, kWh_t^C is the observed energy use in period t in the competition period, and $kWh_t^C\text{^}$ is the predicted energy use in the competition period under the hypothesis that the pilot had no effect on energy use.

Under the assumptions of the model, ES_t^C has a normal distribution with expected value equal to zero. The variance of the energy savings in period t can be obtained by squaring both sides of ES_t^C and taking expectations:

$$\begin{aligned} var(ES_t^C) &= \sigma^2 + \mathbf{X}_t^{C\prime} var(\mathbf{b})\, \mathbf{X}_t^C \\ &= \sigma^2 (1 + \mathbf{X}_t^{C\prime}(X'X)^{-1}\mathbf{X}_t^C) \end{aligned}$$

As σ^2 is not known, we replace it with an estimate s^2.

ES_t^C are the energy savings in only one period of the competition. The project team summed the energy savings in each period to get the competition energy savings ES:

$$ES = \Sigma_{t=1}^{Tc} ES_t^C = \Sigma_{t=1}^{Tc} kWh_t^C - kWh_t^C\text{^}$$

[10] Let e_t be the residual of the regression in period t. s^2 is estimated as the sum of squared residuals divided by T-k, that is, $\Sigma_{t=1}^{T} e_t^2/(T-k)$.

$$= \Sigma_{t=1}^{Tc} [\varepsilon_t^{c} - \mathbf{X}_t^{C\prime}(\mathbf{b} - \beta)]$$

$$= \Sigma_{t=1}^{Tc} \varepsilon_t^{c} - \mathbf{X}^{C\prime}(\mathbf{b} - \beta)$$

where $\mathbf{X}^C = \Sigma_{t=1}^{Tc} \mathbf{X}_t^C$.

Under the null hypothesis, ES has a normal distribution with expected value of zero. Again, the variance can be obtained by squaring both sides and taking expectations. After some algebra, this yields:

$$\text{var(ES)} = \sigma^2 (T_c + \mathbf{X}^{C\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^C)$$

where we have relied on the i.i.d. assumption. Again, as σ^2 is not known, we replace it with an estimate s^2.

The project team used a different formula for the variance of ES. Following the IPMVP (Section B-2.2.2, p. 95-96), the team estimated the variance of the competition period savings as the Mean Square Error (MSE) of the regression*sqrt(T_C) = s^2*T_C. This is an approximation of the actual variance.

TEST STATISTIC

To test the statistical significance of ES, we use the fact that under the null hypothesis ES has a normal distribution with mean of zero and var(ES). The test statistic ES^C/s.e.(ES^C) = $\Sigma_{t=1}^{Tc}$ (kWh_t^C – kWh_t^C^) / s*sqrt($T_c + \mathbf{X}^{C\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^C$) has a *t* distribution with degrees of freedom n-k, where n is the number of observations in the baseline period and k is the number of variables in equation (1) including the constant. Similarly, we can construct a 95% confidence interval, ES^C ± 1.96*s.e.(ES^C), for the true savings. If the test statistic exceeds the critical *t* value or, equivalently, the 95% confidence interval does not include zero, we would reject the null hypothesis of zero savings. We would be reasonably confident that the savings (represented by a significant deviation from the baseline) was not simply chance.

To see how the project team's approximate standard error compares, recall that it equals s*sqrt(n). The exact standard error is s*sqrt ($T_c + X^{C\prime}(X'X)^{-1}X^C$). As $X^{C\prime}(X'X)^{-1}X^C > 0$, the project team approximate standard error is thus an underestimate of the exact standard error.

Project Team Interview Guide

Introduction:

Thank you for taking the time to talk with me today about the Starbucks Pilot program. As you know, Snohomish PUD has asked Cadmus to assess the pilot's approach to measuring energy savings and to gather insights about the pilot's ability to influence energy saving behaviors and other outcomes. We will develop a report that provides the results of our measurement and verification assessment, and that reflects lessons learned from the pilot. Your insights today are essential to our efforts. Your responses are confidential.

We expect this talk to take about an hour. Do you have any questions before we begin?
Is it all right if we record this interview; this is only to help the accuracy of our reporting. YES NO

Please just let me know if I ask you a question that doesn't apply to your role or experience with the pilot and we'll move on.

First I'd like to ask you a little about your background and the roles of those involved in the pilot.

1. What are your primary responsibilities at (organization where the person works)?
2. And what role did you/your organization play with the Starbucks pilot?
3. What were the roles and responsibilities of the other actors involved with the pilot?

Now I'd like to ask you some questions about pilot's intent and design.

4. How did Starbucks get involved with the pilot? What benefit(s) did Starbucks hope to get from participating?
5. What (were) (is your understanding of) the key assumptions – or the theory – underlying the approach used? (Probe: Why were these interventions chosen? – Energy Use Feedback; Energy Saving Tips; Competition'; Establish Norms)
6. What were the goals of the pilot? What was the relative importance of the goals? (Probe: Energy and Non-Energy Benefits)
7. What indicators should be used to measure the pilot's success – that is, to understand how well it met its goals? What information did the pilot gather about those success indicators?

The next questions are about the program's process and flow.

8. Please describe the key steps and activities in the pilot process as it was rolled out. (Probe: How accurately do written materials describe what happened? Are there any significant differences between what's written and actual pilot operation?)
9. How were the stores chosen for participation? (Probe: What requirements did they need to meet?) How were the control group stores chosen?
10. How was their participation communicated to stores? How was their cooperation secured?
11. Once stores were told about the competition, what were the next steps?

Influencing Factors

12. We're trying to understand what features of the pilot design, or outside of the pilot design, influenced store outcomes the most. (Use list of factors below to probe respondents, as needed.)
    a. What do you think were the key factors that made stores more successful?
    b. What do you think were the key factors that limited stores from being more successful?

- Support from upper management
- Conducting kick-off meetings
- Communicating pilot participation goals and expectations
- Having active energy efficiency champions
- Having pre- and post- store surveys
- Having the Lucent dashboard
- Store employee interaction with the Lucent dashboard
- Ensuring involvement of the champions and partners
- Delivery of energy saving tips
- Communication between stores and implementers (e.g., status reports, tips, feedback)
- Use of a friendly competition framework
- Efforts to establish norms
- Past experience of store champions and partners with energy efficiency or sustainability
- Employee turnover
- Needing to meet other organizational priorities

13. What feedback do you have from the store managers, partners, and energy champions about the success of this pilot?
14. What do you think were the greatest strengths of the pilot?
15. What are the most important lessons learned from the pilot?
16. What specific changes would you suggest to make this or similar efforts more effective?
17. How likely do you think it is that the pilot behavior changes are likely to persist? Why or why not?

Let's talk now about coordination and tracking of the pilot program.

18. How effective was coordination among all those involved in implementing the pilot program? What would have improved coordination/communication?
19. Overall, how effective was the data tracking for the pilot? What was tracked? How could tracking have been improved?
20. What, if anything, would improve the accuracy or efficiency of the M&V process used to measure energy savings.
21. Finally, what other insights would be important for us to know about the Starbucks Pilot.

Thank you for your time. We greatly appreciate your input.