

When Failure Can Be a Resounding Success: The Demonstration Project

Arlis Reynolds, Cadmus, Irvine, CA
Keith Miller, National Grid, Waltham, MA
Kimberly Crossman, National Grid, Waltham, MA

ABSTRACT

As energy efficiency markets transform, customers look for new ways to reduce their energy costs, program administrators look for new ways to expand their programs, and entrepreneurs look for new ways to make a profit. Program administrators' inboxes are flooded with announcements of new technologies touting guarantees of large energy savings and high customer satisfaction. In some cases, however, the companies offering these technologies lack data to substantiate their claims or have little insight as to whether and how consumers will use the new technology.

Program administrators perform technology demonstration projects, or tech demos (small-scale tests of specific energy-efficient measures), to assess the savings performance, customer acceptance, and large-scale feasibility of new technologies. This paper uses examples from several years of technology demonstrations to discuss key components of a successful demonstration project, including study designs, sampling strategies, participant recruitment, data collection approaches, and methods for integrating evaluation, measurement, and verification (EM&V) activities into program planning and implementation. For example, the authors discuss how we use our energy laboratory and our own homes to supplement *in situ* data collection, examine the value and reliability of results from small sample sizes, and describe how a failed performance study provided meaningful information for both the program administrator and the manufacturer.

Introduction

To grow programs and achieve deeper savings at customer sites, program administrators must review myriad demand side management (DSM) options and collect sufficient data to design programs that are cost-effective, reliable, and customer-friendly. They then rely on these key data to set program targets and to achieve those targets. Program design typically includes a description of the DSM measure and how it saves energy, eligibility requirements to receive the measure through program activities, estimates of the savings impacts achieved by measure installations, and key implementation practices to ensure both savings and customer satisfaction goals are achieved.

Demonstration projects, also referred to as pilot studies, are used in many types of research as an opportunity to collect critical data and test methods and results before investing in a full-scale program that is relevant and appropriate for their customers. *A pilot study is "A small-scale test of the methods and procedures to be used on a larger scale ..."*¹ *The fundamental purpose of conducting a pilot study is to examine the feasibility of an approach that is intended to ultimately be used in a larger scale study.*²

Energy efficiency program administrators can use demonstration projects to screen new technologies

¹ Porta, M. *A Dictionary of Epidemiology*. 5. Oxford: Oxford University Press. 2008.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3081994/#R13>

² Leon, Andrew C., Ph.D., Lori L. Davis, M.D., and Helena C. Kraemer, Ph.D. "The Role and Interpretation of Pilot Studies in Clinical Research." *Journal of Psychiatric Research* 45(5):626-629. May 2011.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3081994/>

and program designs, examine customer interaction and acceptance, and develop sound savings estimates to set goals and report on programs. Demonstration projects are becoming increasingly important as program administrators continue to look for new solutions to curb energy consumption, manage peak loads, and reduce overall carbon emissions beyond the most common and reliable measures.

This paper discusses how National Grid—a large electric and gas utility serving customers in Massachusetts, New York, and Rhode Island—and other program administrators in the Northeast use pilot studies to test new technologies considered for residential customer offerings or to perform research that informs future program design. In this paper, we discuss examples from five studies that concern various end uses in a residential home (Figure 1).

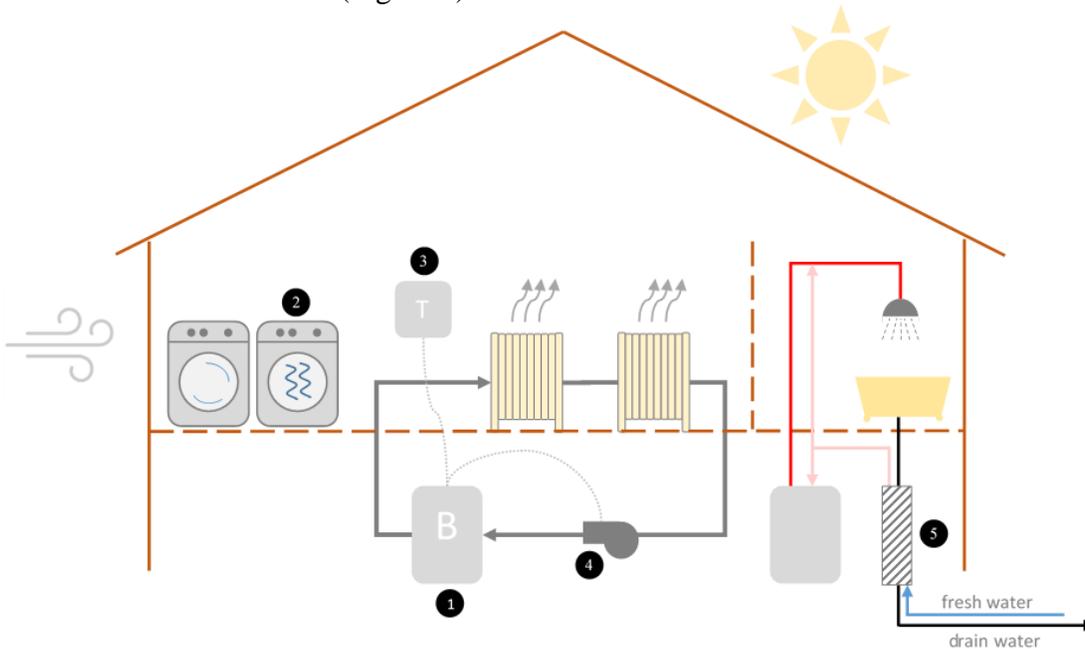


Figure 1. Example Demonstration Projects

The following studies represent a range of methods, budgets, schedules, and level of rigor and the flexibility to design studies based on research priorities, available data, and scheduling needs (numbers correspond to each location in the diagram in Figure 1).

1. **Boiler control** demonstration assessed a new-to-market smart boiler control technology that continuously monitors system parameters to optimize boiler efficiency. The product, a black box (literally) with sensors and a microcomputer, can be retrofitted onto an existing boiler. The manufacturer could not share its proprietary control scheme but claimed that the product would reduce heating energy consumption by 30% on average for residential customers with oil- or gas-fired hydronic heating systems.
2. **Heat pump dryer (HPD)** demonstration assessed two near-market HPDs from different manufacturers. HPDs reduce energy consumption compared to standard dryers (which use electric resistance heating) by recycling the heat and moisture normally disposed through vented exhaust air. Product specifications based on laboratory tests estimate 531 kWh savings per year compared to standard code-compliant dryers.
3. **Smart thermostat** assessment focused on installation and usability characteristics rather than energy performance. The tested thermostats have different product features, installation requirements, and interfaces for customer controls and programming. Our study documented and compared these characteristics to support design of a smart thermostat program. We examined

product features, installation requirements and procedures, and user feedback to rate each model and provide recommendations for National Grid’s thermostat initiative.

4. **Hot water pump** demonstration assessed a load-sensing hot water circulation pump that modulates the flow of hot water with the home’s heating load. In addition to electricity savings from reduced pumping requirements, the manufacturer expects up to 5% reduction in heating fuel consumption due to improved heat transfer efficiency in the boiler.
5. **Drain water heat recovery (DWHR)** demonstration examined the heat recovery performance and potential for a retrofitted shower drain. The DWHR system reduces water heating requirements using a heat exchanger to preheat make-up water with hot waste water. Product suppliers claim that such systems reduce water-heating energy requirements by 35%.

Demonstration Project Process

We first discuss the general process for demonstration projects. Although they can have short timeframes, low budgets, and small samples compared to program evaluations, their scope involves all of the elements of program design, implementation, and evaluation. Figure 2 shows the steps for a typical demonstration project.

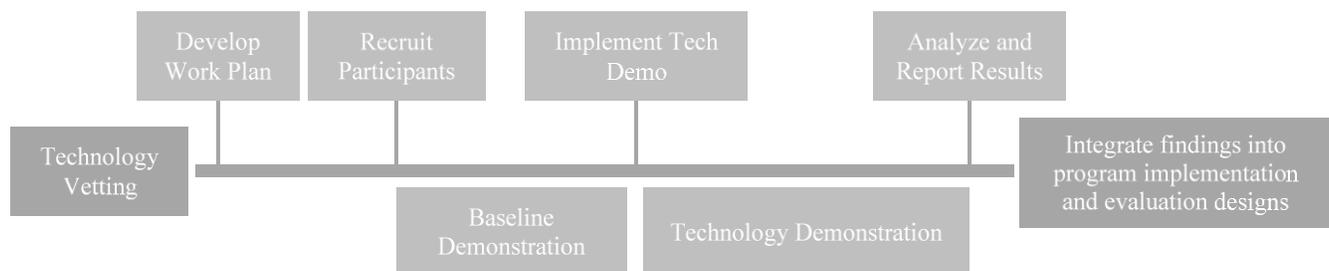


Figure 2. Typical Process for Demonstration Projects

The process is similar to normal program evaluation—work plan development, recruitment, data collection, analysis, and reporting—but it also includes steps that are usually part of program design and implementation activities. The following four common tasks highlight key differences between demonstration projects (or pilot studies) and full-scale program evaluation:

- **Demonstration projects** begin with the **vetting process** to determine the worth of a study. Traditional EM&V activities similarly determine if the level of program savings or spending warrants an evaluation study.
- **Recruitment** for a demonstration project is not limited to a defined participant population. We discuss later that demonstration projects may use many creative recruitment options and may choose from various customer populations. It is important to consider and try to avoid bias in this phase but not critical to eliminate it depending on the project’s subject and priorities.
- **Implementation** is an opportunity to collect baseline data and observe the implementation process. Traditional EM&V is almost always limited to post-installation data collection. Limited knowledge of the implementation process hinders the evaluator’s ability to explain the reasons for unexpected results and provide actionable recommendations.
- **Reporting** should not be the final step for a demonstration project, as it often is for program EM&V studies. Since demonstration projects provide information to determine whether to

include the technology in customer offerings as well as collect key performance data, study results should feed directly into program planning, implementation, and evaluation design.

Although other aspects of demonstration projects appear similar to the traditional EM&V tasks, key differences remain. Through the work plan, sampling, recruiting, and analysis tasks, common EM&V issues such as rigor, bias, and protocol are less critical to providing valuable results. We discuss this issue of relaxed rigor throughout the discussions in this paper.

Vetting

Before we begin planning a research approach for a new DSM product or program design, we must ask, “Does this product or program make sense?” Vetting may involve engineering review (to assess if the technology theory makes sense), a benchmarking review (to assess potential savings or to determine if the claimed savings are reasonable), or a behavioral review (to assess if customers are likely to adopt or reject the product).

Assuming the product passes these screens, each review contributes to the demonstration work plan by identifying data collection needs, availability of data, and potential risks. The vetting process also results in a level of confidence (or skepticism) about the product that informs the level of rigor required to prove the product through demonstration projects. In general—and as we discuss later in this paper—a healthy dose of skepticism is an important element of work plan design.

Work Plan Design for Demonstration Projects

An important difference between demonstration projects and program EM&V is the required level of rigor. Unlike EM&V studies, demonstration projects (and budgets) are not limited to standard protocols or mandatory rigor requirements. Since program EM&V work plans must demonstrate the ability to meet specific confidence and precision levels, meeting these confidence levels within the available budget often dominates the work plan and may in turn limit additional research. EM&V work plans must also follow approved or standard EM&V practices, further limiting the flexibility for relevant program research.

Conversely, demonstration projects are driven by a qualitative confidence. Instead of asking, “Will the results meet 10% precision at 90% confidence,” we ask, “What result would give me confidence to invest in this program” and “What information will be valuable to design the program?” The resulting level of rigor depends on the initial level of comfort with the product, estimated reliability of savings based on experience with comparable products, amount of expected savings, scheduling needs, and available budget.

Demonstration projects can often sacrifice EM&V rigor for efficiency and expediency. Because they are driven by a qualitative confidence, it is not and should not be necessary that demonstration projects meet the same requirements for verifying reported savings to regulators or capacity markets.

But relaxed rigor does not mean results are less valuable. Rather, demonstration projects focus on answering these questions:

- Does the technology work?
- Will customers accept the technology?
- What savings can we realistically expect in an *in situ* environment?
- What are important program design elements to consider for successful implementation?
- What are the barriers that currently inhibit adoption of the technology? (This question could be asked as part of the previous question.)

The demonstration may uncover additional questions we were not aware of and need to answer before operating a full scale program, or it may provide performance data that inform EM&V design and

position the program for a robust full scale evaluation.

Does the Technology Work? Once the theory is vetted, we want to know if, in the context of DSM programs, the technology works to reduce consumption under real-life operating conditions and without negatively impacting customer convenience or comfort. Answering this question typically requires third-party testing of the technology in real-life scenarios.

Boiler control project. We first discussed the control inputs and sequences to identify how the controller saved energy compared to a similar boiler without controls. We designed our data collection plan based on the expected control sequences then installed comprehensive metering on a small sample of boilers to determine if the equipment operated as expected.

A typical evaluation of the boiler control technology would assess energy impacts by examining utility bills that show total energy consumption. For evaluations of heating system retrofits, weatherization measures, or other measures expected to reduce total consumption by more than 10%, comparing pre- and post-retrofit normalized energy consumption is a common approach (IPMVP Option C).³ However, this approach was not sufficient for our boiler pump study for two reasons—our small study population was vulnerable to external influences that skewed savings, and billing data alone would not explain varied or unexpected performance.

We did collect the data to perform the Option C evaluation approach (since it is quick and cheap), but we also collected detailed measurement data on the boilers to examine if the boiler system operated as described by the manufacturer. This additional data collection paid off when our analysis of billing data found no savings or increased energy consumption in 10 of 15 participant homes.

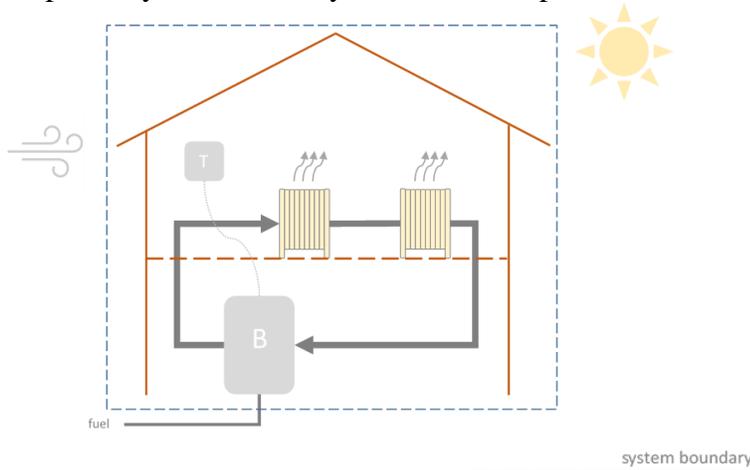
Will customers accept the technology? Verifying that the technology works does not guarantee energy savings in real environments with real customers. For measures that depend on user interaction to achieve savings, the demonstration project should test the measure with customers who fairly represent the expected participant population of a full-scale program. In the residential sector, such measure technologies include smart thermostats (for which savings depend on user schedules and setpoints) and home energy report programs (for which savings depend on occupant response to program data). Similarly, any measure that may impact participant comfort or convenience, or otherwise be noticed by customers, should be tested by representative participants. These technologies include ductless mini-splits (which distribute air differently than the systems they often replace) and HPDs (which typically take longer than electric resistance dryers to achieve the same dryness levels).

The technologies that do not require customer acceptance testing are those with minimal impact outside the energy savings boundary. These include high-efficiency upgrades of like equipment (e.g., replacing an existing refrigerator with a new high-efficiency refrigerator) or measures that reduce the load required to maintain baseline setpoint or comfort levels (e.g., heat recovery).

Figure 3 demonstrates how we analyze system boundaries to determine the likelihood of customers to influence or experience the impacts of a DSM measure.

³ International Performance Measurement & Verification Protocol (IPMVP).

Example 1. System boundary includes user space



Example 2. System boundary excludes user space

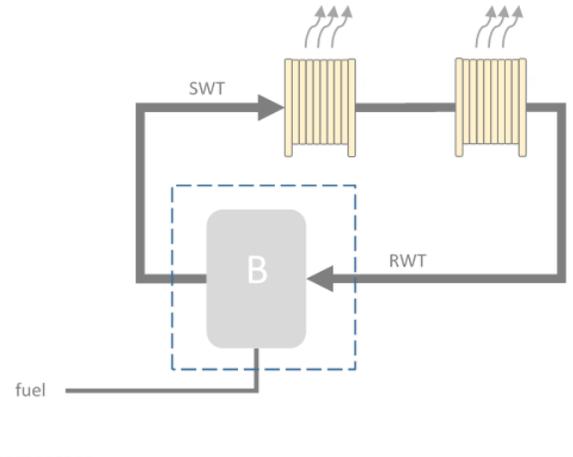


Figure 3. Identifying Customer Interaction

In the house on the left (example 1), a smart thermostat controls the existing boiler to maintain schedules and setpoints determined by the thermostat's optimization program. Compared to a baseline programmable thermostat, the smart thermostat saves energy only if the boiler operates less frequently or produces less heating energy. Providing less heating energy to the space impacts the comfort of home occupants. Therefore, the energy performance of this measure is vulnerable to customer interaction and acceptance. A demonstration project needs to consider this influence and test the measure performance in real environments.

In the house on the right (example 2), a new high-efficiency boiler has been installed in place of a new boiler that only met minimum code requirements. Compared to the code-compliant boiler, energy savings are achieved by the higher combustion efficiency—turning incoming natural gas into usable heat—of the new equipment. As long as the supply water temperature (SWT) setpoint stays the same, the efficiency improvement does not influence the environment outside the system boundary, so the customer does not experience any changes in heating performance. Therefore, this measure would not require customer acceptance testing.

DWHR technology. We demonstrated system performance in a controlled environment then extrapolated results to typical end-user behavior. The DWHR system is an example of a technology whose energy-savings performance is neither influenced nor noticed by the user. Although the total heat recovery potential depends on the volume of hot water used by the homeowner (for which we have data from previous evaluations), the heat transfer performance is influenced only by the temperature of the hot water and inlet water.

To demonstrate the heat recovery performance of the DWHR unit, we installed the system in a colleague's home and monitored system parameters for several months. We used the data to develop a model of system performance and applied that model to typical household water consumption data to estimate annual energy impacts for a typical household.

What savings can we realistically expect in an in situ environment? Demonstration projects should estimate savings using similar techniques as technical reference manuals or program impact evaluations. Since demonstrations typically have smaller sample sizes and shorter metering periods than a program impact evaluation, they often use secondary data to extrapolate results to the anticipated population.

For the HPD study—which involved two months of *in situ* metering at only six customer

households—we estimated average annual energy savings per dryer unit using two key pieces of secondary data. First, we calculated a weighted average unit savings based on the distribution of front-load and top-load washers expected in the population. Since the energy savings performance of the HPD varied with washer configuration, it was important to estimate savings as a weighted average of this distribution. These distribution data were available from a recent statewide residential appliance saturation survey.

Second, we estimated the annual energy savings per unit using the ENERGY STAR[®] appliance calculator standard assumption for average loads per year. Although we had data to estimate annual load for each of our six pilot participants, participant use varied enough to warrant applying this widely used parameter for estimating annual washer use.

What are important program design elements to consider for successful implementation? By piloting the recruitment and installation processes and analyzing technology performance, the demonstration should both provide recommendations to deliver a successful program and define the condition(s) required to achieve predicted savings. For example, these conditions may include customer eligibility requirements, such as age of existing heating equipment, typical occupancy schedules, or use of an existing Wi-Fi system. This information should be clearly documented in the study analysis, and reports should state that the energy savings estimates are based on specific conditions that program participants must also meet to achieve a certain level of savings.

When designing demonstration work plans, we must consider each of these performance categories—acceptance, saving estimation, and program design—and coordinate all activity to maximize useful data and feedback.

Participant Recruitment

After drafting the protocol, the evaluator must find appropriate locations to conduct the demonstration. Earlier, in the customer acceptance discussion, we gave examples of when studies do or do not require *in situ* testing. For measures not influenced by user interaction or acceptance, testing the technology in an *ex situ* (laboratory) or other nonrepresentative user environment may be sufficient. However, for measures that depend on customer interaction or otherwise change the customer's environment, the demonstration should be tested with participants who reasonably represent the target customer population for the potential program. Since demonstration projects often do not have a specific program population from which to recruit, program administrators can be creative with recruiting methods depending on the critical research questions.

There are two important points to consider when identifying potential study participants:

- ***How much do the savings (or other program metrics) depend on the disposition of the customer?*** If the desired results are to establish how the participant uses, interacts with, or responds to the measure, then it is important to engage participants who fairly represent the expected future program population. If the measure performance is immune to user behavior, then bias in participant selection is acceptable.
- ***How burdensome is the data collection on the participant?*** This is especially important because data-intensive demonstrations may require multiple site visits or other onerous data collection tasks. For projects that require *in situ* metering, we may offer large incentives (cash or free equipment such as thermostats) to offset the burden of participation.

For the example demonstration projects in this paper, we used a variety of recruitment techniques based on the specific needs and conditions of the demonstration project. Table 1 characterizes these recruitment techniques and describes how we used these techniques in different demonstration projects.

Table 1. Participant Recruitment Methods for Demonstration Projects

Description	Example
Solicitation from general customer population	
<p>The broadest recruitment effort solicits from the general customer population, which is appropriate if the demo requires minimal pre-screening, if it is important to perform the demonstration with representative participants, or if a PA does not have an existing program relationships with customers. Program administrators can use bill inserts, mass mailings, e-mail distribution lists, cold-calling, or public advertisements. Response rates from the general population are typically very low.</p>	<p>For a smart thermostat pilot program,* the pilot team used the e-mail information from the utility’s customer information system to solicit participants from the general population. The team selected this method because the utility was new to the customer service territory, had limited relationships with customers, and saw the solicitation also as an opportunity to advertise its energy-efficiency programs. We distributed a participation flyer to more than 3,500 e-mail addresses and signed up 30 participants after screening customers for eligibility.</p>
Solicitation from past program participants	
<p>Demonstrations targeting past program participants or requiring a pre-screened population may recruit from program tracking customer lists. This is especially appropriate for demonstrations investigating deeper savings at customer sites. Uptake will be higher than the general population because participants have previously demonstrated interest in energy efficiency and typically have a positive experience with program rebates and reduced energy costs.</p>	<p>For the HWP demonstration, the evaluation team distributed participation flyers to customers who participated in the utility’s boiler replacement program within the last three years. This recruitment modeled the expected program recruitment approach to offer the new product as an add-on to new boiler systems. It also ensured that potential participants had the appropriate equipment and provided participants for which the PA already had verified system data. We distributed flyers by mail and followed up with direct phone calls to recruit 34 participants.</p>
Solicitation from manufacturer	
<p>In collaborative studies, the product manufacturer or supplier may choose to select the demonstration participants. Although this would certainly raise bias flags in program EM&V, it is acceptable for demonstrations as long as the team is comfortable that the selected sites fairly represent future participants. In these cases, the evaluator should be responsible for verifying that the selected participants are appropriate and document any specific conditions on which the demonstration results depend (e.g., baseline equipment more than 10 years old).</p>	<p>For the HPD and boiler control demonstrations, the product manufacturer recruited participants it deemed suitable for the new product. The HPD manufacturer pre-screened participants and homes to ensure the participant was willing to complete intensive surveys for each load of laundry and to confirm the appliance location was suitable in terms of equipment size and weight. The boiler control manufacturer recruited employees with systems identified for savings opportunity and who were open to testing a new system technology. The evaluation team observed all equipment installations and managed all data collection tasks to minimize bias.</p>
Solicitation from friends and colleagues	
<p>When the technology demonstration is the priority and user influence is not a concern, we use the homes of friends and colleagues as convenient test beds. This approach minimizes time, costs, and customer burden. Another advantage is that users already understand the study goals and the importance of quality data. They often contribute thoughtful feedback based on their experiences.</p>	<p>For the HPD and thermostat assessment studies, the utility recruited employees to participate. In exchange for new equipment, the employees agreed to participate in multiple site visits and extensive, manual data logging throughout the four-month demonstration period. For example, for the HPD project, study participants maintained laundry logs, answered questions about the laundry load composition, and weighed each laundry load before (dry) and after (wet) the clothes washer cycle.</p>
Targeted recruitment	
<p>In some cases, demonstration projects may select customers based on specific conditions that facilitate the demonstration. These conditions may be dictated by existing infrastructure (e.g., to collect detailed data) or willingness of the customer to participate in an invasive study. This approach is also appropriate if the planned program will target similarly specific customers. In any case, the evaluator should clearly document those conditions and appropriate baseline conditions that represent a typical program customer.</p>	<p>Although none of the examples discussed in this paper used this method, we have performed targeted recruitment in the past to support technology demonstrations. Motivated to balance the ideal impact evaluation approach—which required hourly baseline data—with low cost, the team has selected buildings with existing whole building time-of-use (TOU) meters and reliable energy management systems to provide trend data for key systems and spaces.</p>

* Cadmus. *Wi-Fi Programmable Thermostat Pilot Program Evaluation*. Prepared for Liberty Utilities. July 2013. <http://www.puc.nh.gov/Regulatory/Docketbk/2012/12-262/LETTERS-MEMOS-TARIFFS/12-262%202013-08-22%20ENGI%20DBA%20LIBERTY%20FILING%20ITS%20PROGRAM%20EVALUATION%20STUDY.PDF>

Implementation

Once customers are recruited, the evaluator can begin gathering data. A key advantage to implementing the demonstration project is the opportunity to collect baseline or pre-retrofit data. As evaluators know, true baseline or pre-installation data are hard to come by, and many evaluated savings estimates (even those based on rigorous evaluation) are still based on modeled baseline consumption. With demonstration projects, we can directly compare before and after operating parameters and clearly describe the measure's impacts on energy performance.

Performing or observing the implementation process is also a valuable opportunity to collect feedback on program design and to identify potential opportunity to collect key EM&V data as part of the implementation process. Evaluators can study these data and make recommendations on eligibility requirements (to screen out installations that may not be successful) or note opportunities for additional data collection during future program implementation processes.

Boiler control and hot water circulator pump demonstrations. We split the heating season to collect system performance data both before and after the retrofit. At the start of the heating season, we performed site visits and installed data loggers to collect pre-retrofit (baseline) data. We then implemented the measure halfway through the heating season and left our data loggers in place to collect post-retrofit data for the remainder of the heating season. By collecting detailed baseline data at the same locations that received retrofits, we could make direct comparisons in system performance and pinpoint the changes caused by the retrofit.

Thermostat assessment study. We examined product features, installation requirements, and usability to help determine which thermostats National Grid should include in its customer offerings. We installed seven different models of smart thermostats in customer homes and then documented the installation process, requirements, timing, and difficulty. Through this, we determined which thermostats would work in most customer homes in the utility's service territory, which thermostats would not work in many homes (due to electrical wiring issues), which thermostats could be installed by any homeowner, which would likely require an HVAC technician, and which might be the simplest to program.

Data Collection

Measure installation marks the beginning of the data collection period. For demonstration projects—which may conclude that a technology does not work—comprehensive data collection is key. Data collection plans are best designed under a state of paranoia, with the assumption that the technology will fail to perform. Then, if the measure does not perform as expected, there are enough data to explain why or at what point the measure is not working or why the analysis shows no savings. Similarly, if the measure demonstrates widely variable performance across test sites, the study data should be able to explain why.

Answering these questions typically requires an intensive metering effort that may be more comprehensive than metering for a normal program impact evaluation. Since demonstration sample designs are typically small, demonstrations can afford this comprehensive metering.

Laboratory, or *ex situ*, testing is a great option to collect data and analyze measure performance in a controlled environment. It also allows stakeholders—program designers and managers, implementers, and evaluators—to practice installation processes, monitor performance, and test data collection equipment. The ability to extrapolate results to a larger, or otherwise different, population gives evaluators more flexibility to use controlled or biased testing sites.

HPD study. We performed a series of controlled tests in a laboratory setting to investigate the influence of various load and operating parameters on dryer performance and energy consumption. By developing a standard test procedure for each dryer and varying only one parameter while holding all other parameters constant, we could examine the impact of single-parameter changes on both the dryer energy performance and the required drying time.

Analysis and Reporting

As with all evaluation activities and regardless of the size of a program, transparency is paramount in analysis and reporting. Both tasks should clearly document the methods of reaching each conclusion, which is especially important for program managers interested in translating demonstration results and data into a full-scale program. The demonstration analysis methods should be comparable to program EM&V analysis methods for estimating energy impacts.

The analysis should also examine the sensitivity of the studied impacts to possible differences between the demonstration and a full-scale program, including the question, “What is the likelihood that savings results will change under different program circumstances?” This information helps the program manager understand the risk in terms of savings targets and determine whether to put additional protections in place such as eligibility requirements for participation, a strict participant screening process, or post-installation follow-up activities (e.g., verification or commissioning).

Reporting should clearly document the metrics, methods, and any assumptions on which the demonstration results are based. If program managers are going to use the results, they need to understand how the results should be applied to a participant or population.

A critical last step in the reporting process is to discuss how the findings, including any assumptions, should influence both program design and future evaluation planning. The evaluator should communicate directly with the program designer and offer support throughout the process to ensure program planning is consistent with the conditions for success that have been presented in the demonstration project’s final report.

Failure as a Positive Result

Consider this definition of failure from the Merriam Webster (online) dictionary: “a failing to perform a duty or expected action.”⁴ For an energy efficiency program that has invested ratepayer funds, promised savings to customers, and reported savings toward an energy efficiency portfolio, this result—failure—would be damaging. However, for a demonstration project whose purpose is to prove or disprove the potential success of a program, failure can be a resounding success. Exposing the consequences of a new products or program approaches in the demonstration stages protects program administrators from adding a risky option to their DSM portfolios or provides important information to redesign implementation strategies. Either result saves program administrators from poor EM&V performance and guides investment of program funds in more promising customer options.

Defining Success

As discussed, the measure of success for a particular demonstration project depends on its specific objectives. Some studies are designed to determine if or how a technology works while other studies may be

⁴ Merriam-Webster, Incorporated. “Failure.” <http://www.merriam-webster.com/dictionary/failure>

designed to develop average savings estimates. However, since demonstration research aims to screen and/or prepare emerging technologies for full-scale programs, we can consider that process complete when our research has met these conditions:

- We have demonstrated that the measure or technology works (typically, to reduce energy consumption or peak loads).
- We understand *how* the measure or technology works and especially any conditions that influence measure performance positively or negatively.
- We have developed savings estimates (or methods for estimating savings) that are reasonable and reliable. We can use benchmarking techniques to determine if savings estimates are reasonable, and we can examine the variability in measure performance across participant sites or test iterations to assess the reliability of the measure performance.
- We have identified the key data required to determine the effectiveness of future technology installations.

In general, a technology that meets these criteria is appropriate for a full-scale program (pending other considerations such as cost-effectiveness, market readiness, etc.). To monitor the performance of a new program based on demonstration research, program administrators should design key parameter data collection into the program implementation plans. Depending on the type of measure and identified data parameters, this real-time data collection may facilitate real-time inspections of measure performance or updates to the original savings estimates. This data collection should also reduce overall costs and timing for full-scale program evaluations.

Recommendations

Program administrators should use demonstration research, or pilot studies, as they explore new DSM solutions for their customers. Lab- or field-based technology demonstrations, small-scale *in situ* studies that test user interaction, and other research provide information critical to successful program design and implementation and facilitate smart investments in the next generation of DSM measures.

Program administrators across the country are investigating technologies and applications that support real-time load management as peak-demand reductions increase in importance compared to energy and carbon reductions. For example, as part of the Reforming the Energy Vision proceedings in New York, program administrators have been asked to design new customer-centric programs that integrate options for energy efficiency and demand response and engage customers as active participants in load management.

When conducting demonstration projects, program administrators should not be bound by protocols or other requirements that may limit creative research. With small budgets and flexible scheduling, they can quickly earn back the costs of demonstration projects through smart investments and robust program implementation and evaluation plans.

References

Cadmus. "Heat Pump Clothes Dryer Technical Demonstration." June 2015. *This study is not publically available.*

Cadmus. "Wi-Fi Thermostat Assessment." April 2015. *This study is not publically available.*

Cadmus. *Wi-Fi Programmable Thermostat Pilot Program Evaluation*. Prepared for Liberty Utilities. July 2013. <http://www.puc.nh.gov/Regulatory/Docketbk/2012/12-262/LETTERS-MEMOS-TARIFFS/12-262%202013-08->

Cadmus. *Impact Evaluation of the 2011-2012 Boiler Reset Control Pilot Program*. August 2012.

Leon, Andrew C., Ph.D., Lori L. Davis, M.D., and Helena C. Kraemer, Ph.D. “The Role and Interpretation of Pilot Studies in Clinical Research.” *Journal of Psychiatric Research* 45(5):626-629. May 2011. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3081994/>

Merriam-Webster, Incorporated. “Failure.” Online dictionary accessed June 2015: <http://www.merriam-webster.com/dictionary/failure>

Porta, M. *A Dictionary of Epidemiology*. 5. Oxford: Oxford University Press. 2008. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3081994/#R13>