National Study of Potential of Smart Thermostats for Energy Efficiency and Demand Response

Jen Robinson, Electric Power Research Institute Ram Narayanamurthy, Electric Power Research Institute Bienvenido Clarin, Electric Power Research Institute Christine Lee, Electric Power Research Institute Pranshu Bansal, University of California, Los Angeles

ABSTRACT

EPRI has implemented a large study of smart thermostats, comprising eight collaborative field implementations and several thousand thermostats in different climates around the United States. Most of the trials include evaluations of energy efficiency and demand response impacts of the thermostats, with some trials also incorporating time-varying rate structures. The trials comprise several different thermostat hardware products as well as various aggregation and optimization platforms. The trial designs were both experimental (e.g., variations on randomized controlled trials, RCTs), and quasi-experimental in nature. One key learning was the difficulty in implementing RCTs when dealing with devices available widely in the marketplace. Preliminary evaluation results suggest smart thermostats are effective in reducing load during peak periods, and customers are generally satisfied with the devices. Furthermore, a preliminary pre-cooling analysis indicates that smart thermostats hold the potential for residential passive energy storage as an alternative to battery storage at a fraction of cost.

Introduction

Connected devices represent an opportunity for utilities to engage with their customers to enable them as grid participants, as well as to provide new choices to enhance customer satisfaction. Much of the activity and market uptake regarding connective devices is centered on smart thermostats—thermostats that customers can control via the internet, including via mobile devices, and which can include enhanced functions like learning algorithms or geo-fencing, as well as offer two-way communication capabilities. Figure 1 contains EPRI's typology of the different categories of thermostats—manual, programmable, and "smart". Furthermore, inherent to their two-way communications capabilities, smart thermostats generate new data streams regarding set point preferences, HVAC equipment operation, indoor and outdoor temperature, and more. These data streams themselves represent new opportunities—to understand utility customers and their heating and cooling needs and behavior, to understand premise-specific building envelopes and their potential for improvement, to understand customer-specific HVAC equipment operation characteristics—the list goes on.

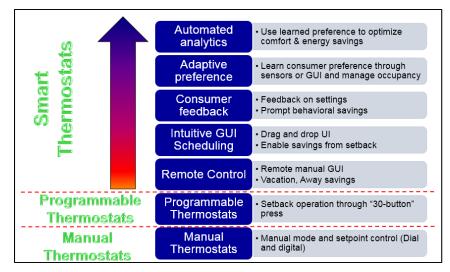


Figure 1. Thermostat Categorization Including Smart Thermostat Variations

The excitement regarding these new opportunities, as well as the need for more research to critically assess any potential benefits, led to the creation of EPRI's multi-utility smart thermostat collaborative project. To date, the project includes 17 utility collaborators, eight of which are fielding pilots or detailed secondary research analyses to assess the potential energy efficiency (EE) or demand response (DR) impacts of smart thermostats. The project's parallel research efforts also include technical assessments of commercially available smart thermostats, thermostat-level data analytics, and broader industry stakeholder engagement.

In this paper we review some of the pilots, including the learnings from attempted implementations of randomized controlled trials, some preliminary results, as well as discussion on the potential use of smart thermostats for passive thermal storage in buildings.

Pilot overviews

To date the EPRI smart thermostat collaborative project has involved the design of four US-based utility pilots: one in the Southeast, two in the Midwest, and one in the Southwest. Analysis is underway for all four pilots, as well as a fifth in the northeastern U.S. (EPRI did not contribute to the pilot design in this case). Table 1 is a summary of five of the project's pilots for which analysis is underway. At the time of publication, implementation was also underway for two additional pilots encompassing several thermostat products and ~1,400 devices.

Pilot	Pilot Focus	Approx. No. of Devices and Brand by Treatment Group	Pilot Design Type
South- eastern utility (#1)	DR, summer and winter	200 smart thermostat brand 1	Quasi-experimental: within subjects and matching
Mid- western utility (#2)	DR, summer	 220 smart thermostat brand 2 210 direct load control switches on outdoor unit (cellular + Wi-Fi) 	Experimental: some randomized DR events Quasi-experimental: within subjects
South- western utility (#3)	EE, load shifting, & bill savings, summer; customers are on a time-of-use (TOU) rate	140 smart thermostat brand 3	Quasi-experimental: recruited control group Experimental back- up: randomized encouragement design (RED)
Large Mid- western utility (#4)	EE and DR, summer; EE, winter	 630 smart thermostat brand 4 600 smart thermostat brand 5 (Plus ~680 control group customers that received either thermostat after the summer test period) 	Experimental: RCT using recruit and delay Quasi-experimental back-up: matching
North- eastern utility (#5)	EE and DR, summer; EE, winter	 1,000 smart thermostat brand 1 1,500 smart thermostat brand 2 	Quasi-experimental: matching + waitlisted control group

Table 1. Summaries of Some EPRI Collaborative Smart Thermostat Pilots

Pilot designs—aspirations and reality

The pilot design process for Pilots #1 through #4 began by considering the use of the most rigorous design approaches, particularly for those pilots whose aim was to assess energy efficiency (EE) impacts—that is, the impacts of the thermostats on overall energy use. Specifically, we began by considering the implementation of an experimental design—that is, where the allocation of customers to treatment and control groups is determined using randomization. For device-based pilots that require customers to volunteer, these can be

implemented via randomized controlled trials (RCTs) employing "recruit and deny" or "recruit and delay" approaches.¹

When the more rigorous methods were not possible—and they often were not—we continued down the pilot design hierarchy, building off past EPRI work (EPRI, 2013), to find the next-best alternatives in the "quasi-experimental" realm—that is, pilot designs where randomization was not used in any manner, or where it was originally intended, but it could not be properly implemented. Figure 2 illustrates the pilot designs of each. In the end, of the four pilots, one could be considered an experimental design (a randomized controlled trial employing a recruit and delay approach), and the remaining three were quasi-experimental, although two of the three incorporated some degree of randomization.

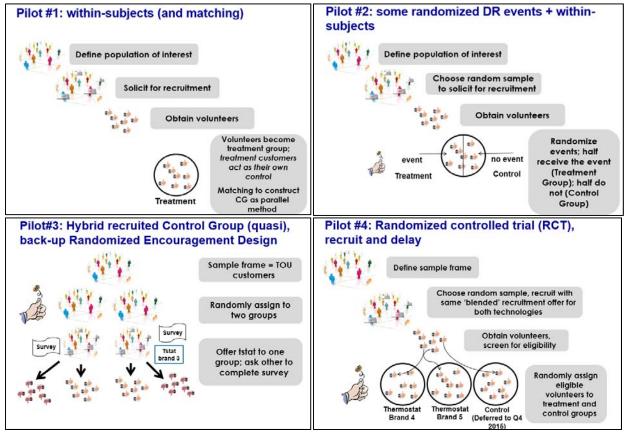


Figure 2. Pilot Design Structures-experimental and quasi-experimental

Even when not always successful in trying to implement the pilots as experiments (that is, using random assignment), several lessons were learned along the way regarding the feasibility of using experiments for device-based pilots:

¹ In RCTs employing "recruit and deny", customers are recruited to the pilot as they would be with any utility program, but volunteers are then randomly assigned to either the Treatment or the Control Group. Control Group customers do not receive the intervention (e.g., the thermostat), and thus recruitment messaging needs to be clear about that upfront—for example, by framing the pilot offer as a lottery. "Recruit and delay" designs are similar, but Control Group customers receive the intervention after the end of the test period. Several resources explain pilot methods in detail, including SEE Action Network, and EPRI, 2013.

- Randomized experiments take time and resources to design and implement properly. Compared to standard program recruiting practices, their implementation may seem onerous and even bizarre ("... recruit a customer, and then not give them a thermostat?"). Ensure the project team understands the value and tradeoffs of implementing an experiment versus a quasi-experiment, and make the pilot design decisions together to ensure that the level of rigor required is justified by the utility's circumstances (for example, will pilot results be used to justify expensive business decisions?). Then, lay out the facts and obtain buy-in from senior management upfront.
- Possible customer experience concerns regarding recruit and deny/delay designs can be mitigated through transparency to the customer regarding the nature of the pilot, and with appropriate message framing. Recruit and deny approaches work well when framed as lotteries or a "chance to win." If possible, the fact that not all customers will receive a device (right away or at all) should be reiterated to customers at some point directly before enrollment. A backup plan for potentially dissatisfied customers should also be in place.
- Often recruitment can be outsourced to third parties, so it is crucial to remain on top of every detail of that implementation, both from a customer experience perspective and to ensure that the randomization is performed properly. Months of work in designing an experiment can be quickly undone with the smallest missed implementation detail.
- While social and mass media campaigns are perfectly appropriate for marketing standard programs, they should not be used to recruit customers to a pilot designed as a randomized controlled trial with recruit and deny or delay. One exception may be when there is confidence that recruitment targets can be reached within a relatively short period, although as the evidence from multiple smart thermostat pilots suggests, achieving recruitment goals can be difficult, and multiple recruitment waves are often necessary. Sending multiple recruitment requests via mass or social media channels can send mixed signals to customers that may have already volunteered but were randomized to the Control Group. Even when pilot messaging is transparent and customers are told upfront they may not receive a thermostat, being turned down and then continuing to see multiple recruitment appeals would lead to a poor customer experience.
- For recruit and deny/delay pilots, it is advisable to make the Control Group disproportionately smaller than the Treatment Group, if possible. This will have the effect of raising the minimum energy or demand savings that the pilot will be able to detect, which may not always be acceptable. However, if this is acceptable, this means that fewer people need to be denied or delayed.
- Having a parallel pilot design structure can be helpful as a backup plan when it is unknown whether a particular design will have enough power to detect effects.
- When quasi-experimental pilot designs are necessary, if possible, solicit survey responses from the Control Group using instruments that are identical in nature to the Treatment Group survey because it may help in making the Treatment and Control Groups more similar, on average. Another benefit is that having survey data from Control Group customers opens the door to multiple quasi-experimental approaches, such as propensity score matching. Although it would be a matter of opinion whether this would be a more

• accurate approach than using the Control Group as is, it allows analysts the option to, for example, use multiple approaches in parallel in order to choose the most conservative outcome.

Some preliminary results

Complete analyses are expected to be public by Q2 2016, although a snapshot of some results can be shared in the meantime:

DR savings

Preliminary results from Pilot #1's winter DR events suggest average hourly peak reductions in the 16 to 19% range for 130 customers, just over a quarter of whom have electrically heated homes. Table 2 includes results achieved through two different (parallel) methods—within subjects and using matching—to develop a frame of reference for comparison.

Table 2. Pilot #1's PRELIMINARY Impact Analysis Results (Winter 2013/2014) (result at a 95% level of significance)

Analysis Approach		Average Hourly	Snapback ²	
		Reduction for 130 customers		(10 a.m12 p.m.)
		(7-10 a.m. Central)		
1	Within-subjects (no	-16%*	0.28 kW	Not significant
	Control Group)	-10%	(0.85 kWh per event)	(+5%)
2	Matched Control Group	-19%*	0.31 kW	Not significant
	Comparison	-1970	(0.93 kWh per event)	(+3%)

* At a 99% level of significance

Customer opinions

Customer reaction to the smart thermostats is generally positive overall. Figure 3 includes some satisfaction results for Pilot #2, and Figure 4 demonstrates that most customers clearly prefer the smart thermostat.

² Snapback is defined as the consumption impact after the control period ends.

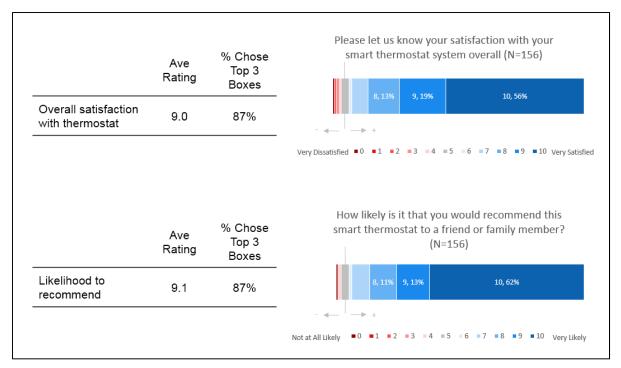


Figure 3. Pilot #2: Overall satisfaction with smart thermostat, likelihood to recommend to others.

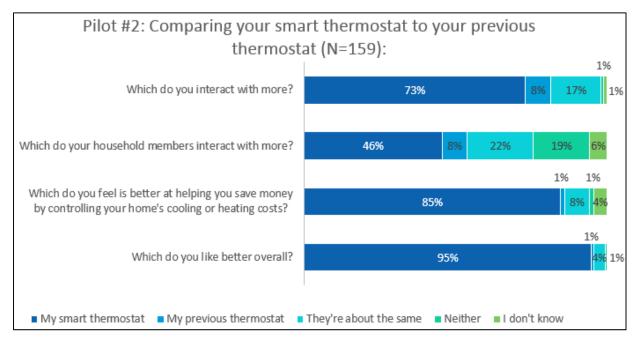


Figure 4. Pilot #2: Preference compared to previous thermostat.

Thermostat programming behavior

More than half of Pilot #2 customers had a programmable thermostat originally—both the cohort that received the smart thermostat as part of the pilot, and the cohort that received the switch (but no new thermostat). Both cohorts that had a programmable thermostat prior to the pilot were

asked whether they programmed their thermostats, both before their new pilot devices were installed, and after. The results in Figure 5 suggest that a significantly larger proportion of customers programmed their thermostat after their smart thermostat was installed. Some caveats:

- The smart thermostats were installed with default settings programmed, which could have contributed to the increase in programmed thermostats (indeed, that is the intention of default settings).
- Slightly different wording regarding the "hold" question in the pre- and post-installation surveys may explain some of changes—this may be evidenced by the change in the programming behavior of the switch customers, which was not expected. Recall the switch customer cohort had no new thermostat installed, only a switch placed on their outdoor unit.
- The results are based on self-reported data, which can biased.

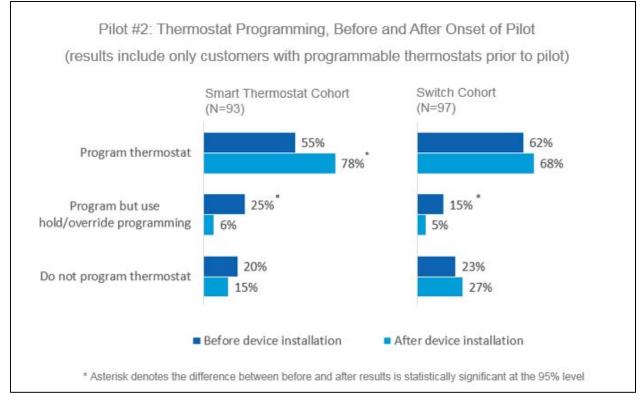


Figure 5. Changes in the number of thermostat programmed (comparing pre- and post-installation survey results).

While Figure 5 alone is not enough to prove that the smart thermostat *caused* an increase in the number of programmed thermostats, pending survey analyses from the other pilots may shed additional light on this question.

Use case: thermostats for passive thermal storage

Smart thermostats may represent an opportunity to achieve passive storage benefits at a relatively low cost in existing residential buildings. The ability to send mass signals to adjust

thermostat set points, and in particular the ability to "pre-cool" or "pre-warm" before a peak period, is foundational to enabling passive thermal storage in residential buildings. Precooling/pre-warming relies on the thermal mass inherently available in a premise, which can be used to store thermal energy, essentially using the space as a thermal battery.

The available storage capacity of buildings can be difficult to quantify as it depends on several variables, including building insulation, solar heat gain, building occupancy patterns and customer comfort preferences. As a result, analysis of passive storage capabilities will need to account for these differences.

During the summer of 2015, as part of Pilot #2, several DR events were held that included a pre-cooling phase prior to the peak period. On September 2, a load shed event was conducted that spanned a total of 5 hours (1-6 p.m. CDT), with the first two hours consisting of pre-cooling with a -3 °F set point offset, followed by the main event from 3 to 6 p.m. that involved a +3°F set point offset. This event was randomly assigned to the smart thermostat customers such that only a portion (~60%) were subjected to the load shed event. The other pilot participants were not sent the load shed event, and therefore in effect created a control group for the impact analysis.

The utility does not have advanced metering infrastructure (AMI), so hourly premiselevel data were not available. Instead the data collected by the smart thermostats was used to conduct the passive thermal storage analysis. Compressor run-time data from the thermostat was coupled with HVAC system information to estimate HVAC-level energy usage.³ The load profiles of the Treatment and Control Group customers are depicted in the left side of Figure 6. In addition, the load profile of the Treatment Group customers on a similar weather day (August 15th) was also recorded as a Baseline Day, and this is depicted in the right graph in Figure 6. Energy usage for each group is tabulated in Table 3.

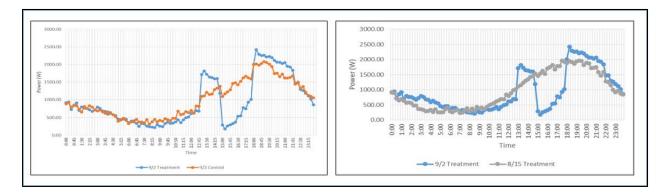


Figure 6. Power vs. Time Plots comparing Treatment and Control Groups on the DR Event Day (left), and the Treatment Group on the Event Day and Baseline Days (right).

Table 3. Average Energy and Power of Treatment and Control Groups per Household for DR Event Day (09/02/2015) and Baseline Day (08/15/2015)

³ The method used for estimating HVAC-level energy consumption from smart thermostat data and HVAC data collected at thermostat installation is detailed in two EPRI reports: EPRI, 2015 and EPRI, 2016.

Time Period	Treatment Group		Control Group	
	Event Day	Baseline	Event Day	Baseline
		Day		Day
All Day (kWh)	18.6	23.4	24.2	20.6
Noon -10 p.m. (kWh)	12.5	16.2	15.2	13.8
3 - 6 p.m. (kWh)	2.0	5.6	4.8	5.6
Peak Demand (kW)	1.0 kW	2.0 kW	2.0 kW	1.7 kW
Average Outside Air Temperature (°F)	81.6	78.3	81.6	78.3
Average Indoor Air Temperature (°F)	76.7	76.0	76.1	75.8

Table 3 shows a decrease in peak period (3 - 6 p.m.) energy use and demand of the Treatment Group on the Event Day compared to the Baseline Day. Adjusting for the Control Group, this is roughly equivalent to the average passive storage capacity attributed to the homes' thermal mass. These data are for one single demand response event (and Baseline Day) in a particular Midwestern utility's service territory, and further pre-cooling demand response events are still needed to confirm overall passive storage opportunities. Also, storage capacity benefits may not directly transferable to other utility service territories, and capacity results may vary based on building characteristics in a utility service territory such as building age, HVAC efficiency, weather conditions and customer preferences.

Coupling and/or comparing passive thermal storage methods with customer-sited electric storage enabled by batteries should also be assessed. With the cost of smart thermostats currently anywhere from \$75-\$300USD, this technology potentially offers similar load shifting capabilities at a fraction of the cost of residential electrical energy storage systems. Tradeoffs between cost-effectiveness for utility and customer, plus reliability and response time should be considered when making the assessment to couple and/or compare these two storage solutions.⁴ It is important to note, that although this home showed passive thermal energy storage potential, moving a customer set point past his or her comfort band may lead to forfeiture in this potential and poor customer comfort. Utilities need to consider these tradeoffs when designing utility deployments.

Conclusion

It is expected that consumer interest in smart thermostats will continue to grow (EPRI, 2015c), and in so doing, so will their potential to be used as tools to enable customer engagement and grid participation. Preliminary evidence from the EPRI pilots, as well as from other studies, suggests they are effective in enabling demand response, and they are generally likely by customers. A growing body of evidence suggests they may enable overall energy savings as well (Brannan, 2015), although the evidence is mixed—further analysis of their EE potential is required, and it is the aim of the EPRI pilots to contribute to this. Finally, new use cases, such as the potential use of smart thermostat for cost-effective passive thermal storage, will continue be explored to examine new opportunities for services and savings.

⁴ See EPRI, 2015b for a brief on the economics of residential energy storage.

References

- Brannan, D. 2015. Getting Smarter? Evidence of Savings from the Nest Thermostat. Presentation from the 2015 Behavior, Energy & Climate Change Conference, Sacramento, California, October 19, 2015.
- EPRI, 2016. Assessing Data Made Available from Smart Thermostats and its Potential Applications for Utility Program M&V (forthcoming). EPRI, Palo Alto, CA: 2016. 3002008956.
- EPRI, 2015. *Real-Time M&V: Opportunities and Barriers*. EPRI, Palo Alto, CA: 2015. 3002005819.
- EPRI, 2015b. *EPRI Quick Insights: The Economics of Residential Energy Storage*. EPRI, Palo Alto, CA: 2015.
- EPRI, 2015c. You Ask, We Investigate Webcast Series, 2015, Number 2: Bring Your Own Device Program Approaches. EPRI, Palo Alto, CA: 2015. 3002006184.
- EPRI, 2013. *Quantifying the Impacts of Time-Based Rates, Enabling Technology, and Other Treatments in Consumer Behavior Studies: Protocols and Guidelines.* EPRI, Palo Alto, CA: 2013. 3002000282.
- State and Local Energy Efficiency (SEE) Action Network. 2012. Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations. Prepared by A. Todd, E. Stuart, S. Schiller, and C. Goldman, Lawrence Berkeley National Laboratory.