ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

Facilitating Energy Savings through Enhanced Usability of Thermostats

Alan Meier[1], Cecilia Aragon[1], Therese Peffer[2], Daniel Perry[2], Marco Pritoni[3]

[1]Lawrence Berkeley National Laboratory, Berkeley, California, USA
[2]University of California, Berkeley, California, USA
[3]University of California, Davis, California, USA

Environmental Energy
Technologies Division

2011

Published in the Proceedings of *ECEEE Summer Study on Energy Efficiency*. ECEEE 2011 Summer Study. Belambra Presqu'île de Giens, France: European Council for an Energy-Efficient Economy.

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Technology Development, Building Technologies Program, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Facilitating energy savings through enhanced usability of thermostats

Alan Meier
Lawrence Berkeley National Laboratory
1 Cyclotron Road, Berkeley
California 94720 USA
akmeier@lbl.gov

Therese Peffer
California Institute for Energy and Environment
University of California, Berkeley
California 94720 USA
therese.peffer@gmail.com

Marco Pritoni
Department of Mechanical Engineering
University of California, Davis
California 95616 USA
mpritoni@lbl.gov

Cecilia Aragon
Lawrence Berkeley National Laboratory
1 Cyclotron Road, Berkeley
California 94720 USA
cragon@lbl.gov

Daniel Perry
School of Information Sciences
University of California, Berkeley
California 94720 USA
dperry@lbl.gov

Keywords

control and monitoring devices, cooling, buildings, heat controls, user perspective, usability, interface

Abstract

Residential thermostats play a key role in controlling heating and cooling systems. Occupants often find the controls of programmable thermostats confusing, sometimes leading to higher heating consumption than when the buildings are controlled manually. A high degree of usability is vital to a programmable thermostat's effectiveness because, unlike a more efficient heating system, occupants must engage in specific actions after installation to obtain energy savings.

We developed a procedure for measuring the usability of thermostats and tested this methodology with 31 subjects on five thermostats. The procedure requires first identifying representative tasks associated with the device and then testing the subjects' ability to accomplish those tasks. The procedure was able to demonstrate the subjects' wide ability to accomplish tasks and the influence of a device's usability on success rates. A metric based on the time to accomplish the tasks and the fraction of subjects actually completing the tasks captured the key aspects of each thermostat's usability.

The procedure was recently adopted by the Energy Star Program for its thermostat specification. The approach appears suitable for quantifying usability of controls in other products, such as heat pump water heaters and commercial lighting.

Introduction: Thermostats control a large amount of energy

Residential thermostats play a key role in controlling heating and cooling systems, especially for single-family homes. In a typical home, the thermostat often controls over 30 % of a home's total energy consumption. As such, they control a large amount of national energy use in North America and Europe. In the United States for example, thermostats control about 9 % of national energy use (Peffer et al. 2012). With such a large amount of energy at play, it is essential to understand the thermostat's technology, and the way in which the occupants interact with them.

In 1995, the Energy Star Program established technical specifications for "energy saving" programmable thermostats (U.S. Environmental Protection Agency 2010). Many building codes and government programs now require installation of programmable thermostats because of their assumed energy savings. However, there have been few careful studies of the energy savings attributable to these thermostats. Several recent field studies have found no significant savings in households equipped with programmable thermostats compared to households with manual thermostats (Cross and Judd 1997; Haiad et al. 2004; Nevius and Pigg 2000; Michelle Shipworth et al. 2010). Two other studies found that homes relying on programmable thermostats actually consumed more energy than those where the occupants set the thermostats manually (Sachs 2004), especially in homes with heat pumps (Bouchelle, Parker, and Anello 2000). Other studies in the United States, the United Kingdom and Finland suggested that the thermostats were overly complex and that consumers were unable to operate them in a way that obtained energy savings compared to manually operated thermostats. These findings coincide with a wealth of anecdotal evidence that occupants are baffled by programmable thermostats. As a result, Energy Star terminated the thermostat endorsement program in 2009 and decided that any future endorsement program must include specifications for minimum levels of usability. We describe below a method to measure a thermostat's usability. The method can be applied to other products where the user interface is critical to the product's performance.

The energy savings from programmable thermostats differ from installation of more efficient refrigerators or heating systems. The purchase of an energy-efficient product results in energy savings without any further action by the consumer. In contrast, programmable thermostats require further consumer action to save energy: the occupants must program the thermostats to shift to lower temperatures and re-set them when schedules change. The literature survey and responses to our own surveys (Meier et al. 2011) suggest that many consumers believe that simply purchasing and installing an "energy-saving" programmable thermostat will automatically result in lower energy use. More intuitive designs, interfaces, and procedures will *facilitate* greater energy savings for all segments of the population. This paper first describes earlier research into usability of thermostats and then proposes a method for quantifying usability in future thermostats.

The poor usability of thermostats has been mostly ignored

The *usability* of thermostats has been the subject of relatively little research even though it is a popular complaint and topic for anecdotes. To be sure, thermostat manufacturers have undertaken research into the effectiveness of their designs, but the results have been mostly confined to proprietary reports. Manufacturers consider any insights gained through their usability studies to be a competitive advantage. Furthermore, manufacturers tend to focus on their own products rather than examining the effectiveness of different classifications of the devices.

A survey of the literature broadly dealing with thermostats and usability was undertaken prior to this research effort (Meier et al. 2010). Researchers have periodically commented about usability problems associated with thermostats both when specifically examining thermostats or in the course of other research. Table 1 summarizes the usability problems identified in the literature. Surprisingly few comments have been made over

Table 1. Usability problems associated with programmable thermostats identified in the literature adapted from (Meier et al. 2010).

Programmable Thermostats Complaints/Issues
PTs are too complicated to use
Buttons/fonts are too small
Abbreviations and terminology are hard-to-understand; lights and symbols are confusing
The positioning of interface elements is illogical
PTs are positioned in an inaccessible location
Setting the thermostat is troublesome
It is difficult to set time and date
PTs give poor feedback on programming
PTs are not attractive to use

(Note: "PT" = Programmable Thermostat)

the past twenty years, especially compared to investigations of other components in heating and cooling systems.

An important concept is the mental model assumed by thermostat users. Kempton (Kempton 1986) used ethnographic methods to interview occupants and building supervisors to derive insights. For example, many occupants treated thermostats more like a valve rather than a switch. Thus, the occupants expected heat to be delivered faster when they set higher target temperatures. This led to energy-wasteful operating outcomes because indoor temperatures would overshoot desired temperatures. Our own research indicates that this remains a popular mental model (Pritoni et al. 2011).

Problems with thermostats have been observed in several countries. In Finland, Karjalainen (2009) made qualitative and quantitative surveys on thermostat use in homes and offices. He concluded that many people had misconceptions about how thermostats and their heating systems actually operate (such as treating the thermostat as a valve rather than a switch) and that they found thermostats too complicated to use with confidence.

In Britain, Rathouse and Young (Rathouse and Young 2004), conducted six focus groups to investigate issues in use of heating controls. They found that many consumers had serious misconceptions about the relationships between thermostat settings, comfort, and energy use. Based on the users' experiences and complaints, Rathouse and Young formulated recommendations for manufacturers and installers, including a proposal that manufacturers offer products of varying complexity to suit different needs. Conversations with manufacturers of cooling and heating devices in Japan and Korea indicate that usability is a problem with their equipment, too.

Consumer associations have evaluated thermostats in Europe and North America. Usability was typically one of several factors considered in the overall ratings. These evaluations generally took place in conditions where usability problems would be minimized. For example, when *Consumer Reports* (2007) evaluated fifteen thermostats, the tests were conducted in a well-illuminated room, by highly-trained panellists comfortably seated at a table (a situation rarely encountered in homes). Even then, the panellists found some of the thermostats difficult to use. Consumer magazines in other countries, notably Germany (Stiftung Warentest 2008) and Sweden (Råd och Rön 2003) have also reviewed thermostats. Both reviews included usability as a consideration but only in a qualitative sense. The British Energy-Savings Trust endorses certain programmable thermostats but its Central Heating System Specifications (Energy Saving Trust 2008) do not include usability criteria for "programmers" (e.g, programmable thermostats).

We concluded from our survey that a procedure to measure and quantify usability would contribute to identifying thermostat designs that aid consumers in finding energy-saving settings, or at least not create an unneeded barrier. We therefore created a test procedure to quantify the usability of thermostats.

A method to quantify usability

A TEST PROCEDURE BASED ON TASKS

We investigated the feasibility of quantifying usability of thermostats through controlled interactions between people and thermostats. Ideally, the test method should resemble an energy test procedure, that is, be clearly defined, and have quantifiable, repeatable, results. These measurements of usability could then be used to establish a "usability score" which would allow manufacturers, consumers, and regulatory agencies to rank thermostats and establish minimum criteria for usability.

The measurement method involves two steps:

1. Define representative tasks to be accomplished with the device;
2. Measure people's ability to perform those tasks under controlled conditions using defined metrics.

The first step in measuring usability is defining the most common tasks associated with the space heating features of the thermostat. A "task" might be as simple as ascertaining the status of the thermostat; for example, "Identify the temperature the thermostat is set to reach." Alternatively, a task might involve changing the operation, such as, "Program the temperature to be 22 °C on Tuesday evenings at 7 PM." We compiled a list of tasks by studying the operating manuals and observing and interviewing users. We also solicited recommendations from manufacturers. Finally, we required that all standard programmable thermostats had the capabilities of accepting that task. From this long list, we selected six tasks that a typical user would need to understand in order to effectively operate the programmable thermostat. The six tasks were:

Task 1: Turn the thermostat from "off" to "heat."

Task 2: Set the correct time on the thermostat's internal clock.

Task 3: Identify the temperature the device is set to reach.

Task 4: Identify the temperature that the thermostat is set to reach for Thursday at 9:00 PM.

Task 5: Put the thermostat in "hold" or "vacation" modes to keep the same temperature while occupants are away.

Task 6: Program a schedule and temperature preferences for Monday through Friday.

The above tasks are clearly defined and can be easily explained to test subjects. Successful operation of a programmable thermostat requires proficiency in other tasks but these are representative; in other words, if users can perform these tasks, then they can use the most important features of the thermostat. The same approach could be applied to other sorts of controls, such as lights or heat pump water heaters.

Our review of the literature found no earlier attempts to quantitatively measure usability of thermostats or similar devices. Therefore we were unsure which measurement (or metric) would best capture the sense of usability. We sought to observe in detail and record six different aspects of usability with which we might construct metrics of usability. The following aspects were collected for each subject during each test:

- success or failure in accomplishing the task;
- elapsed time to accomplish the task;
- number of times buttons were pushed (or other actions);
- sequence of actions;
- hesitations; and
- verbal comments.

A good metric will successfully differentiate between skilled and unskilled users and between intuitive and opaque interfaces. A metric could consist of one or a combination of these aspects. By collecting all of this information, we could search for the best metric. We recorded the sessions with a video camera; this way we were able to convert the data collected on a video record and quantify the characteristics listed above.

Our initial goal was to determine the viability of the task-based methodology and the identification of the best metric. Did the test procedure generate a significant range in the metrics? When applied to different thermostats, did the test procedure generate a significant range in a metric? Finally, was one metric superior to others?

DETAILS OF EXPERIMENT

Five programmable thermostats were selected for testing. Three were primarily controlled through a touchscreen, one through buttons, and one employed a web-based interface. The tests were conducted at a usability laboratory. The laboratory set-up was very simple (see Figure 1). A video camera recorded each test in the vicinity around the thermostat (so the subject's face was not captured). The camera captured images similar to that shown in Figure 2.

The subjects were mostly recruited through online classified postings to sections for "creative gigs" and "labor gigs" in the San Francisco Bay area. Two were recruited from a similar posting to a university e-mail list. Participants came from varied occupations and backgrounds, including construction workers, business managers, non-profit staff, maintenance workers and students. Participants were asked to rate their previous experience with programmable thermostats. Seventeen people reported their experience level as "low," eight as "moderate" and five reported having "no experience with programmable thermostats" (one participant gave no response). In the end, 31 participants were recruited, ranging in age from 18 to 65. Nine were female. All participants were given a small financial incentive for taking part in the study.

Each subject was tested on two thermostats. Each test consisted of six tasks. Altogether 62 tests were performed, consisting of 372 tasks. The subjects did not receive any training prior to being tested; however, an operating manual was placed on a table next to the thermostat which they could consult if they wished.

Results: Metrics of usability

A wide range of usability was observed. For example, in Task 1, that is, switch the thermostat from off to set heat, 26 % of the subjects were unable to accomplish the task at all. Figure 3 displays the completion fraction for each thermostat. For Thermostat A, all subjects successfully completed the task. In contrast, only 50 % of the subjects using Thermostat E completed the task. Similar results occurred for the other tasks (but are not presented here).

In Figure 4 the elapsed time for each subject to accomplish (or fail to accomplish) Task 1 is plotted for each thermostat. Some subjects were able to accomplish Task 1 in less than ten

Figure 1. Laboratory set-up for measuring the usability of a thermostat.

Figure 2. Still image from a video of a subject performing Task 1.

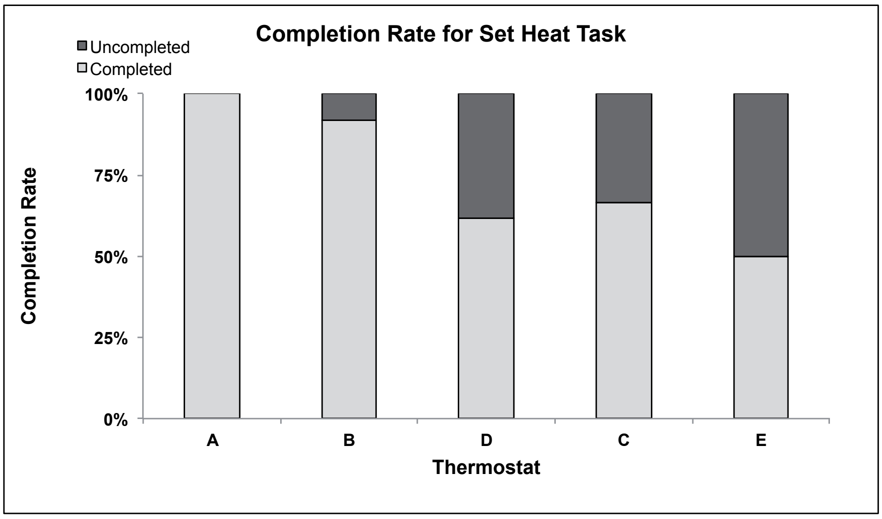

Figure 3. Fraction of subjects that successfully completed the Set Heat Task.

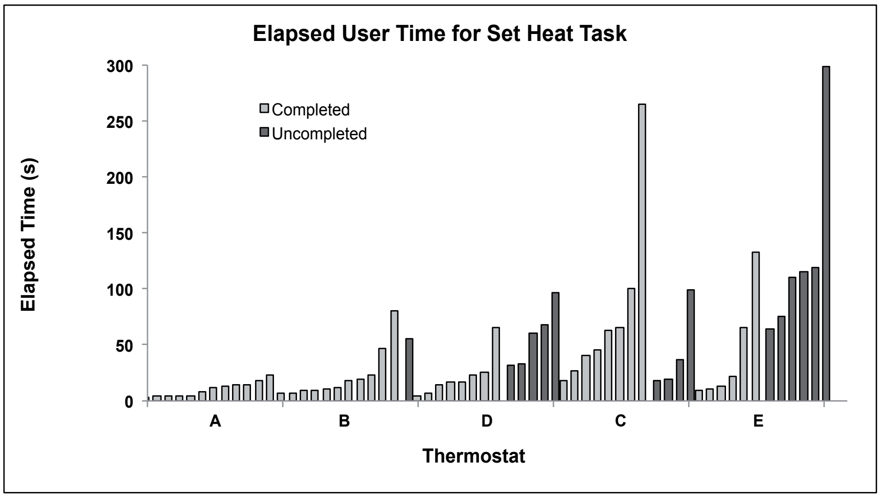

Figure 4. Elapsed times for subjects to perform the Set Heat Task, including times for those who were not successful.

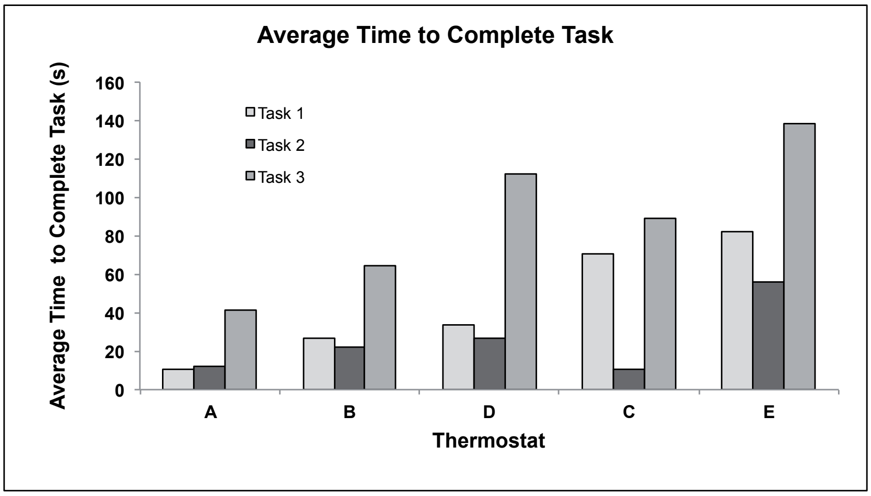

Figure 5. Average elapsed times for subjects to complete Tasks 1, 2, and 3.

seconds. Most subjects were able to accomplish the task in less than 30 seconds; however, over 30 % of the subjects required 31–120 seconds. (Note that two minutes can feel like a very long time when standing in front of a thermostat trying to switch on the heat.) These results indicate the range in the subjects' ability to perform this task. Similar results occurred for the other tasks (but are not presented here).

The time for subjects to accomplish the task varied widely for the same thermostat, too. For Thermostat C, one person successfully switched on the heat in 20 seconds, while another required 260 seconds. The remaining times were evenly distributed between the minimum and maximum times. Wide ranges in elapsed time occurred for Thermostats C and E but much less so for A, B, and D. Elapsed times for completion and success rates do not have a clear correlation. Compare Thermostat A, which had 100 % success rate and very low elapsed times, to Thermostat E, which had a low success rate and a wide range in elapsed times.

Thermostats D and E had hinged covers which concealed some of the controls. Many subjects were unable to open the cover or did not recognize that a cover existed, resulting in more failures to complete. This illustrates how small design differences can have large impacts on successful operation of a device. Note that this task captures a subject's first encounter with the thermostat; the results could change once he or she becomes familiar with its operation. However, a continuing lack of familiarity–or "forgettability"–may be a reasonable assumption if occupants rarely interact with their units.

The results shown in Figure 3 and Figure 4 (and other results not shown here) demonstrated that the methodology produced a wide range of measured abilities of the subjects to perform the task.

A second requirement for the task-based methodology is the ability to quantitatively differentiate levels of usability among thermostat interfaces. Figure 4 displays the range in elapsed time to completion for accomplishing Task 1 with the five thermostats. The figure demonstrates that the task-based methodology and the metric achieved clear differentiation among the thermostats. The average time to accomplish Task 1 for Thermostat E was roughly eight times longer than for Thermostat A. Thermostats A and B were clearly superior (for this task) because the subjects were able to accomplish the task quickly and nearly all of the subjects successfully completed the task. In contrast, the subjects accomplished Task 1 on Thermostat D relatively slowly and a significant fraction were unable to complete it at all.

The results were similar for other tasks. Figure 5 shows the average elapsed times for Tasks 1, 2, and 3. A wide range in average completion time was observed in all three tasks. The ranking of thermostats changed slightly depending on the task but, in general, a model with long average completion times for one task had long completion times for other tasks. Note that results for Tasks 4, 5, and 6 are still being evaluated but appear to be similar to the first three tasks presented here.

Figure 4 and Figure 5 used the metric of elapsed time for comparison of subjects and thermostats. Other metrics were investigated, including, the percentage of subjects that completed the task, the number of button pushes, and the ratio of observed button pushes divided by the minimum required. We found that all of the metrics produced sufficient ranges in results and all of the metrics generated the same ranking of usability for almost every task. The consistency of these results point to the robustness of the overall task-based approach to measuring usability.

An improved usability metric

Average elapsed time for completion is an attractive metric because it is simple to understand and measure; however, elapsed time is misleading since the metric ignores those who fail to complete the task. We therefore explored a hybrid metric, combining both elapsed time to complete and successful completion of the task. We also sought to develop a metric that would be easier for manufacturers, regulators, and other stakeholders to interpret and compare. A common drawback of many usability metrics is that the value of the metric is unbounded and varies from task to task. This creates confusion; it is not obvious what value of a metric is "good" and the metric cannot be compared on an absolute scale from one task or device to another. For programmable thermostats, stakeholders need a single measure of usability to facilitate consumer understanding and to create an absolute scale of usability that is not dependent on arbitrary task length.

In order to create such a metric, we chose a variant of the logistic function,

$$\frac{2}{1+e^{x}}$$

that maps $[0,\infty)$ to the interval $[1,0)$. In other words, a shorter time for completion is mapped to a value close to 1, and a longer time is mapped to a value closer to 0.

We also needed to account for success rates on a per-trial basis (where a task "trial" is a single instance of a participant performing a task on a thermostat model, also sometimes called a "task observation"), rather than averaging over all trials of a given task. In order to accomplish this, we incorporated the task completion or success rate variable, *s*, directly into our primary equation, which we called the "*M*" statistic. The "*M*" statistic is calculated as follows on a per-trial basis:

$$M_i = \frac{2s}{1+e^{x}}$$

where

$x = t/k$

$$s = \begin{cases} 0, \text{ if subject failed to complete task} \\ 1, \text{ if subject completed task} \end{cases}$$

t = time for subject to complete the task (seconds)

k = 50 (empirically determined constant)

Note that *M* will always be normalized between 0 and 1. The success rate variable, *s*, also always falls between 0 and 1. It can be a binary variable (where $s = 1$ if the task is completed and 0 otherwise), have multiple values for partial success (e.g. if the task has several subparts that can be completed successfully), or be a continuous variable that measures percentage of task completion.

The *M*-statistic combines time to complete the task with success of the trial in an intuitive manner: if the task is not completed so that $s = 0$, the value of the *M*-statistic is 0. Intuitively, this means that if the task was not completed, it should not matter how long the user spent attempting it; it is still a failure. If, on the other hand, the task is completed successfully, then the time on task weighs into the *M*-statistic. For example, a shorter task duration will yield a higher value of *M*, a longer task duration will yield a lower value of *M*, and an uncompleted task will set $M = 0$.

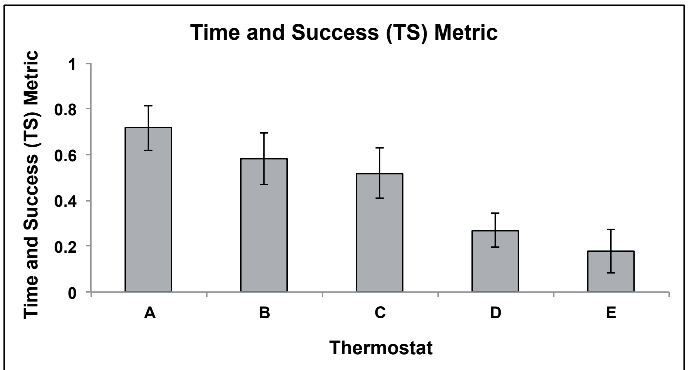

Figure 6. Combined time and success metric for Tasks 1, 2, and 3.

We found that a metric combining time to completion and success to complete was the most practical (which we called the "time and success metric"). The results for the three tasks, using the M-statistic (and $k_1 = 50$) are shown in Figure 6. An analysis of variance (ANOVA) showed that, for the time and success metric, the effect of thermostat model on usability was significant at $p < .01$.

Figure 6 shows mean values, along with error bars at the 95 % confidence level. Both of the concepts, time to completion and success to complete, are easy to understand. Furthermore, they are easy to measure in a laboratory with relatively simple equipment. These features make the time and success metric an attractive metric for quantifying usability.

Discussion

Evaluating usability of products is commonplace; however, most evaluations address usability of one-off items, such as controls in airline cockpits and power plants or organization of web pages. The typical procedure is to compare one version against an improved version. To our knowledge, this is the first quantitative usability test developed for mass-produced products. The test is intended to generate a value or score to demonstrate that the product exceeds a minimum level of usability.

These preliminary results suggest that it is possible to quantitatively evaluate the usability of thermostats. These results also suggest that a usability score, based on a combination of tasks, will be a meaningful indicator of overall usability. The results are promising but further research is still needed in the areas mentioned below:

- We tested several metrics of usability and found that they all gave essentially the same rankings. We found that one metric that combined time to complete, ability to complete, and a logistic form made best use of the data and best captured usability. Moreover, the data for this metric are relatively simple to collect and are unambiguous. Other kinds of tests, such as assessing the number of hesitations or corrections, can provide insights into usability. Biological measures of stress and frustration also deserve investigation.
- How many tasks need to be created to adequately represent overall usability? Every test procedure is a trade-off between realism, cost, and repeatability. We arbitrarily selected six because it seemed about right.
- To what extent should the tests take into account subjects learning and becoming familiar with interfaces? The subjects' performance might change dramatically if the tests were immediately repeated. On the other hand, this may not be a realistic situation if occupants perform these tasks only a few times per year.
- How many people should be on a user test panel and how should they be selected? These questions require guidance from both statisticians and policymakers. On the statistical side, we need large enough test panels to attain satisfactory confidence in the results. Policymakers need to decide to what extent elderly, handicapped, colour-blind, and non-English speakers are included.

- Repeatability is a key requirement for any test procedure. We have not yet confirmed that the test results can be duplicated in other laboratories.
- Can repeatability be improved by testing subjects on a "reference" interface in addition to the product under test? This approach would lessen distortions caused by non-representative sampling.
- Does the test procedure stifle innovation? Thermostats are undergoing rapid changes in both technologies and requirements. For example, can this test accommodate voice commands or visual cues?

Energy Star is addressing many of these issues (U.S. Environmental Protection Agency 2010) because it intends to include a usability criterion in its next specification for programmable thermostats (which it calls "climate control devices"). To our knowledge, this Energy Star specification is the first application of a quantitative usability requirement for the controls of a device.

APPLICATION OF THIS APPROACH TO OTHER PRODUCTS

The task-based approach, as well as the metric, can be applied to other devices where poor usability may impede energy-saving behaviour. Heat pump water heaters require sophisticated controls so as to ensure maximum efficiency while meeting hot water needs. At the same time, incorrect settings of these controls can lead to greatly increased energy consumption. The likelihood of incorrect settings is high because controls are confusing and occupants are not familiar with this new device. Figure 7 shows the controls for three commercially-available heat pump water heaters.

Advanced lighting controls in commercial buildings pose similar usability problems. Occupants will probably default to a lights-on mode (if only for safety considerations) if a control is too complicated to operate, which will result in unnecessarily high electricity use.

Conclusions

We developed a laboratory test procedure to measure usability of different thermostat interfaces. The test is based on subjects performing a set of representative tasks needed to effectively operate the thermostat. Our methodology successfully differentiated people's abilities to perform tasks and successfully differentiated the usability of different thermostat interfaces. These results assured us that the methodology is robust and will allow manufacturers and regulators to quantify a thermostat's usability.

Energy Star recently proposed to use a slightly modified version of this methodology in its new specification for Climate Control Devices. Stakeholders are evaluating it and will soon prepare responses. In addition, an independent test laboratory is applying the procedure to the same thermostats in order to confirm repeatability and to identify weaknesses in the test procedure. In September 2011, Energy Star will release a specification for Climate Control Devices incorporating minimum usability levels based on this research.

Minimum levels of usability could easily be added as another criterion to minimum energy performance standards (MEPS)

Figure 7. Controls for three residential heat pump water heaters (source: Electric Power Research Institute).

for appliances. These minimum usability levels would ensure that consumers would obtain a technically efficient product while ensuring that operators could easily achieve the lowest possible energy in a simple and transparent manner.

Quantifying usability is an attractive area for cooperative international research. There remain numerous methodological aspects to examine, such as size and composition of the test panel, repeatability, and use of a reference interface, Research to date has not revealed any unique national or cultural aspects that would prevent sharing research results. No country has established usability requirements for the controls of energy-using appliances so this research would not interfere with regulatory activities already underway.

A more usable thermostat does not guarantee lower energy use but a less usable thermostat certainly presents an obstacle to reducing energy use. We assume that improving the usability of thermostats will *facilitate* energy-saving behaviour. Other kinds of programs still need to educate consumers about appropriate settings and behaviours.

References

Bouchelle, Matthew P., Danny S. Parker, and Michael T. Anello. 2000. Factors Influencing Space Heat and Heat Pump Efficiency from a Large-Scale Residential Monitoring Study. In *Proceedings of the 2000 ACEEE Summer Study on Energy Efficiency in Buildings*. Pacific Grove (Calif.): American Council for An Energy-Efficient Economy.

Consumer Reports. 2007. Programmable Thermostats Lab Test – Some make saving easier. http://www.consumerreports.org/cro/appliances/heating-cooling-and-air/thermostats/thermostats-10-07/overview/therm-ov.htm.

Cross, D., and D. Judd. 1997. Automatic Setback Thermostats: Measure Persistence and Customer Behavior. In *Proceedings of the International Energy Program Evaluation Conference*. Chicago: IEPEC.

Energy Saving Trust. 2008. *Central heating system specifications (CHeSS) Year 2008*. London UK: Energy Saving Trust, May. http://www.defra.gov.uk/sustainable/government/documents/chess-2008.pdf.

Haiad, Carlos, John Peterson, Paul Reeves, and John Hirsch. 2004. *Programmable Thermostats Installed into Residential Buildings: Predicting Energy Savings Using Occupant Behavior & Simulation*. Rosemead: Southern California Edison.

Karjalainen, Sami. 2009. Thermal comfort and use of thermostats in Finnish homes and offices. *Building and Environment* 44, no. 6: 1237-1245.

Kempton, Willett. 1986. Two theories of home heat control. *Cognitive Science* 10, no. 1: 75-90.

Meier, Alan, Cecilia Aragon, Therese Peffer, Daniel Perry, and Marco Pritoni. 2011. Usability of Residential Thermostats: Preliminary Investigations. *Building and Environment* (in press).

Meier, Alan, Cecilia Aragon, Therese Peffer, and Marco Pritoni. 2010. *Thermostat Interface and Usability: A Survey*. Berkeley (Calif.): Lawrence Berkeley National Laboratory, September.

Nevius, MJ, and S Pigg. 2000. Programmable Thermostats That Go Berserk: Taking a Social Perspective on Space Heating in Wisconsin. In *Proceedings of the 2000 ACEEE Summer Study on Energy Efficiency in Buildings*, 8.233-8.244. Pacific Grove.

Peffer, Therese, Marco Pritoni, Alan Meier, Cecilia Aragon, and Daniel Perry. 2012. How People Use Thermostats in Homes: A Literature Review. *Building and Environment (submitted for publication)*.

Pritoni, Marco, Alan Meier, Therese Peffer, Daniel Perry, and Cecilia Aragon. 2011. How Turkers Use Thermostats. *In Preparation*.

Råd och Rön. 2003. Test: Värmestyrsystem. *Råd och Rön*.

Rathouse, Kathryn, and Bruce Young. 2004. *RPDH15: Use of Domestic Heating Controls*. Watford: Building Research Establishment (UK).

Sachs, Harvey. 2004. *Programmable Thermostats*. Washington, D.C.: American Council for an Energy Efficient Economy.

Shipworth, Michelle, Steven K. Firth, Michael I. Gentry, Andrew J. Wright, David T. Shipworth, and Kevin J. Lomas. 2010. Central heating thermostat settings and timing: building demographics. *Building Research & Information* 38, no. 1 (April 12): 50 - 69.

Stiftung Warentest. 2008. Heizkörperthermostate - Ausgewählt, geprüft, bewertet. *Test*, May.

U.S. Environmental Protection Agency. 2010. ENERGY STAR® Residential Climate Controls Draft Specification Framework –Performance-Based Usability Requirements. November 30. http://www.energystar.gov/ia/partners/prod_development/new_specs/downloads/climate_controls/Residential_CC_Draft_Specification_Framework.pdf.

Acknowledgements

The authors thank Gari Kloss for her enthusiastic assistance in all aspects of this research. This work was supported by the Office of Energy Efficiency and Renewable Energy, Building Technologies Program, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.