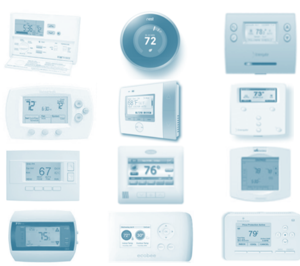
SMUD's Communicating Thermostat Usability Study



A simultaneous, multi-user, paired comparison test of communicating thermostats for task efficiency, preference, and perceived usefulness of advanced features

July 2013

(DRAFT)

Herter Energy
RESEARCH SOLUTIONS

Prepared by: Herter Energy Research Solutions, Inc.
2201 Francisco Drive, Suite 140-120
El Dorado Hills, California
www.HerterEnergy.com

Authors: Karen Herter, Ph.D.
Yevgeniya Okuneva, Statistician

Prepared for: Sacramento Municipal Utility District
Sacramento, California

Project Manager: Melinda Brown

SMUD Contract No: 4500071792

Full report available online at
www.HerterEnergy.com/pages/publications.html

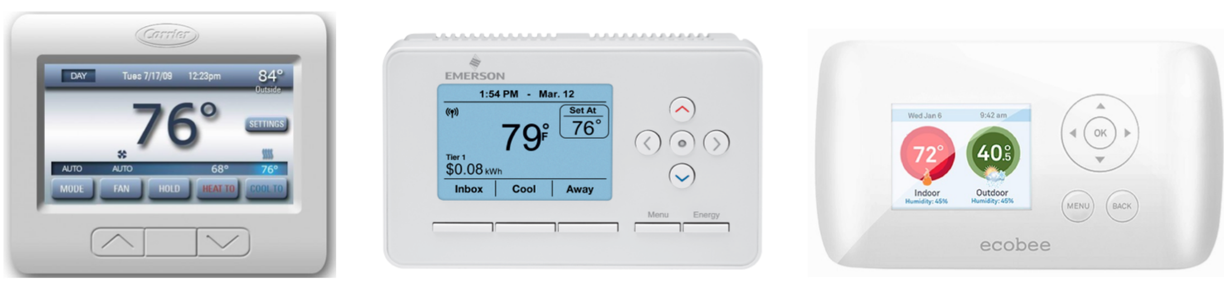
EXECUTIVE SUMMARY

The goal of this study was to better understand the features that contribute to ease of use and preference for thermostats, and in particular, communicating thermostats. Data was collected during a 3-day lab study, during which 12 thermostats were each tested by between 26 and 28 participants, evenly distributed by age, education, income, home ownership, and gender.

Efficiency scores were calculated from time-on-task values derived from videos of each of the 326 thermostat tests. *Preference* scores were based on data collected in surveys indicating the preferred thermostat of each participant. Results of linear regression models that incorporated these efficiency and preference scores along with indicators for thermostat features and participant characteristics indicated the following:

> **Preference.** Preference scores were significantly higher for thermostats with color displays and high overall feel and sound ratings. The three most preferred thermostats were the Carrier ComfortChoice Touch, Emerson Smart Energy, and Ecobee Smart Si.
>
> **Efficiency**. Efficiency scores, based on time required to complete standard tasks, were significantly higher for thermostats with larger screens and higher ease of use ratings. The three most efficient thermostats were the Carrier ComfortChoice Touch, Honeywell FocusPro (one of two non-communicating units tested), and Emerson Smart Energy.

FIGURE 1. TOP SCORING COMMUNICATING THERMOSTATS



Carrier ComfortChoice Touch   Emerson Smart Energy   Ecobee Smart Si

Based on these and other findings of this study, the research team recommends the following for future utility programs that involve thermostats:

1. Establish minimum threshold usability scores as a prerequisite for purchase.
2. Conduct usability tests for all thermostat models being considered for programs, to determine whether they meet the minimum threshold usability scores.
3. Provide extra training for renters and the elderly, who took significantly longer to complete common tasks than did younger participants and homeowners.

TABLE OF CONTENTS

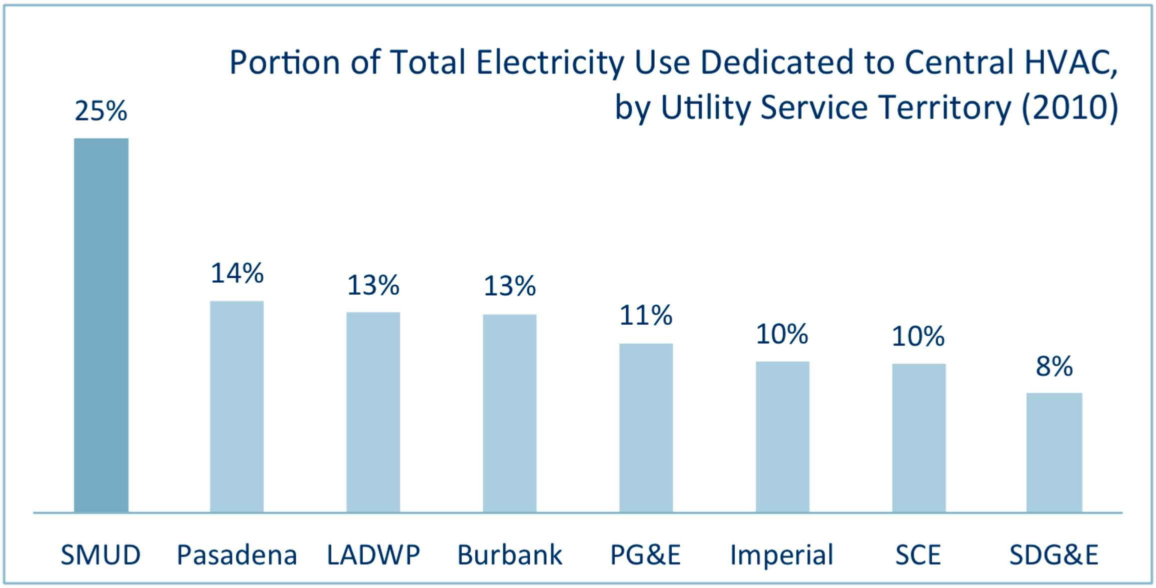
List of Figures

List of Tables

1 BACKGROUND AND GOALS

The thermostat is an unassuming yet ubiquitous device that plays an increasingly large role in residential electricity use. In SMUD homes, thermostats control one-quarter of all electricity consumption – a fraction unmatched by any of the other major utilities in California (Figure 2). On the hottest summer days, residential air-conditioning is responsible for about one-third of SMUD's total 3,000-megawatt peak demand.

FIGURE 2. RESIDENTIAL HVAC ELECTRICITY USE IN CALIFORNIA



Source: California Energy Commission, 2009.

1.1 THERMOSTAT STANDARDS

Until recently, little has been done to take advantage of the energy savings opportunities inherent in thermostatic controls. Early efforts focused on the use of programmable setpoint schedules for reducing HVAC use when occupants were regularly away or asleep. Since the early 1980's, the California Energy Commission's Title 24 building standards have required that thermostats have four such programmable setpoints, designed for *Wake*, *Day*, *Evening*, and *Sleep* periods.

In 1995, the U.S. Energy Star program borrowed these programmable setpoint specifications for their voluntary thermostat certification program. About a decade later, however, Energy Star rescinded the thermostat certification program, citing several studies showing that the programming features were not being used properly, or at all, and that the promised savings had not materialized (Figure 3). Since then, Energy Star has been working with vendors and researchers to devise a new set of specifications. The current proposed specifications require communications to allow "3rd party developers to enable access to the product's full range of communication and remote control capabilities." (Energy Star, 2012)[1]

Figure 3. Programmable thermostat studies

Location	Organization	Year	Homes	Conclusions
Connecticut	Connecticut National Gas Corp.	1996	100	No significant change
Wisconsin	Energy Center of Wisconsin	1999	299	No significant change
Florida	Florida Solar Energy Center	2000	150	No savings, some increases
Northwest	Bonneville Power/PNNL	2001	150	No significant change
California	Southern California Edison	2004	N/A	Some savings, some increases

Source: Gunshinan 2007.

To date, the California Energy Commission has not followed Energy Star's lead in repealing the original 4-setpoint requirements; however, like Energy Star, California is pursuing a standards update that includes remote communications. The Commission's first attempt at setting a standard for "Programmable Communicating Thermostats" (or PCTs) was abruptly shelved in early 2008 when the media caught wind of plans to require emergency-based remote control of thermostats by utilities (New York Times 2008).

After removing the controversial requirement and renaming the devices *Occupant Controlled Smart Thermostats* (OCSTs) to emphasize the update, the Commission adopted OCSTs into California's 2013 building standards. (See Appendix B.) Communication and messaging must use standards-based protocols such as IP or ZigBee, OpenADR or Smart Energy Profile, and communications hardware may be built-in or removable.

Starting in 2014, OCSTs will become mandatory in nonresidential buildings, except where an energy management and control system (EMCS) fulfills the same functionality. At the same time, OCSTs will be a compliance option for residential new construction, as a trade-off for part of the solar-ready dedicated roof area requirement.

In summary, then, it appears that a transformation of thermostat functionality is imminent. The sudden shift towards communicating thermostat standards opens the potential for new communications-based functionality to be provided on a mass scale. The question remains, however, whether these new standards are an improvement on the previous standards, i.e. whether the new thermostats will be used in a way that actually helps customers use less energy. While it's too soon pass judgment on the far end of the communications path, we can say with some certainty that these new standards will not effect energy savings if customers

[1] See Appendix A for the current proposed Energy Star specification for thermostats.

don't like or can't figure out how to use the new thermostats. Our goal in this study, then, is to investigate and compare the usability and likeability of 10 of the newest thermostats on the market alongside two non-communicating thermostats commonly found in Sacramento homes.

1.2 An Overview of Usability Testing

The National Institute of Standards and Technology (NIST) defines usability as "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." (NIST 2001)

Some of the basic elements of usability testing are as follows (Rubin and Chisnell 2008):

- Development of research questions rather than hypotheses
- Use of a representative sample of end users, which may or may not be randomly chosen
- Representation of the actual environment
- Observation of end users who either use or review a representation of the product
- Controlled interviewing and probing of the participants by the test facilitator
- Collection of quantitative and qualitative performance and preference measures
- Recommendation of improvements to the design of the product

This study made use of a specific implementation of usability test called a *comparison test*, where the purpose is to explicitly compare two or more products. The basic methodology involves side-by-side comparison of two or more different product designs. Performance and preference data are collected for each product and the results are compared. Comparison tests are typically used to establish which design is easier to use or learn, and to better understand the advantages and disadvantages of different designs (Rubin and Chisnell 2008).

The unique contribution of user testing is that it exposes what people actually *do*, as opposed to what they *say* – or say they will do. Paired-comparison testing was chosen for this study to elicit more honest critical feedback than would single-unit testing. With paired comparison testing, users are forced to choose one product over the other, and are given the opportunity to explain in more detail precisely what they liked about one product in relation to what was *not as good* about the other (Enerson 2012).

Although this study compares several instances of thermostats – twelve to be specific – the focus of the evaluation was to consider individual features, with the expectation that the best overall design may not be present, but rather a conceptual hybrid combination of the many feature options.

2 Study Overview

This document describes the implementation and evaluation of a simultaneous multi-user, multi-device comparison test of thermostats available for purchase in 2013. The testing was conducted to enable comparison of the short-term or "walk-up" usability of twelve different thermostats, meaning that participants were not provided with user manuals or coached in any way prior to their interaction with the thermostats.

2.1 Goals and Objectives

The primary goal of this study was to assess the features and functions of a variety of communicating thermostats to determine which characteristics might be recommended or required in specifications for thermostats promoted by or implemented for future programs at SMUD. The objectives of the study were to:

- Calculate and compare usability metrics for a sample of new thermostats
- Determine preferences for communicating thermostat features
- Identify specific design concerns, particularly for those thermostats involved in current or planned programs at SMUD

2.2 Research questions

- How do performance efficiency metrics compare between products?
- How do satisfaction metrics compare between products?
- How do participants rate the advanced features they reviewed?
- What features are most helpful to users in completing common tasks?
- What flaws prevent users from completing common tasks?
- How do products rank in order of which is chosen most often as the favorite?

2.3 Approach

- Test 10 advanced and 2 standard thermostats under controlled lab conditions
    - Devise a list of common tasks for each participant to perform
    - Video record participant attempts to complete the task list
    - Conduct discussion sessions to gain further qualitative insights
- Review the video recording of tasks
    - Record done or not done for each task (success)
    - Record start and end times for each task (time-on-task)
- Establish baseline usability and satisfaction levels
    - Establish efficiency and satisfaction metrics
    - Calculate efficiency metrics for each task
    - Collect satisfaction metrics through surveys
    - Conduct statistical analysis of metrics

2.4 Schedule

Month	Task
January	Develop research plan
	Procure equipment
February	Design and construct testing cubicles
	Draft scripts, surveys and other documentation
	Internal beta testing of process
March	Finalize testing process, scripts and documentation
	Perform trial run through of entire test with SMUD recruits
April	Recruit test participants
	Usability test
	Input survey data
May	Review video for time-on-task and success rates
	Calculate efficiency, satisfaction and overall usability ratings
	Draft Report
June	Final Report
	Final Presentation at SMUD

3 Study Approach

3.1 Overview

In the spring of 2013, SMUD conducted a study to collect qualitative and quantitative data for use in comparing user interactions with and reactions to 2 common and 10 communicating thermostats. Each participant was asked to test two of the twelve thermostats to enable comparison of the *walk-up* usability of the thermostats. Participants were not provided with user manuals or coached in any way prior to the usability testing.

Test sessions involved an introductory discussion with the facilitator, the first thermostat test and survey, a discussion session, the second thermostat test and survey, another discussion session, and end. Interactions were video recorded as participants performed a realistic set of tasks using each thermostat. After each thermostat test, participants filled out a survey designed to collect ratings for thermostat usability, look, feel and sound. At the end of the each thermostat test, a discussion session was held and recorded.

After three days of testing with participants, survey data, time on task measurements, and success rates were recorded and analyzed, indicating statistically significant differences in the ease of use metric and clear user preferences for certain thermostat features.

The following sections describe this approach and the data analysis results in greater detail.

3.2 Roles During Testing

Following are the roles of the people involved during the three days of lab testing.

Facility Staff

Facility staff members were present for the entirety of the testing to:

- Direct participants as needed
- Provide assistance with internet connectivity issues
- Provide miscellaneous items that aided in testing (e.g. batteries, staplers, pens, etc.)
- Provide food and beverages for the research team and participants

Facilitator

The facilitator was present for the entirety of each test session to:

- Provide an overview of the study and the purpose of usability testing to participants
- Indicate start and stop times for testing
- Conduct group discussion sessions
- Respond to non-technical requests for assistance

Participants

For each of the two thermostats tested, the participant's role was to:

- Attempt to complete a set of representative task scenarios as efficiently as possible
- Fill out a post-test questionnaire
- Participate in a discussion session to provide honest opinions regarding the usability and likability of the thermostats and supporting applications

Technical Support Staff

The technical support person was present for the entirety of each test session to:

- Monitor recording equipment
- Resolve technical problems with thermostats or other equipment

Principal Investigator

The principal investigator was present for the entirety of the testing to:

- Observe testing and take notes
- Direct facilitator and technical support staff as needed
- Address issues that could not be resolved by the technical support staff or facilitator
- Answer viewer questions

3.3 Participant Sample

A total of 180 residential SMUD customers were recruited for participation in the thermostat usability study based on a sample size power analysis showing the need for a minimum of 24 participant tests per thermostat. (See Appendix C). Using the recruitment script provided in Appendix D, about 15 participants were recruited for each of 12 cells defined by 6 age categories and 2 education categories. Of the 180 recruited participants, 163 attended the usability testing, as summarized in Table 1.

Table 1. Number of participants that attended the testing, by age and education

Age	Year of Birth	<4 years college	4+ years college	Total
18 – 28	*1985 – 1994*	16	12	**28**
28 – 38	*1975 – 1984*	12	13	**25**
38 – 48	*1965 – 1974*	14	11	**25**
48 – 58	*1955 – 1964*	14	17	**31**
58 – 68	*1945 – 1954*	15	13	**28**
68 +	*- 1944*	14	12	**26**
Totals		**85**	**78**	**163**

Demographic Distribution

An effort was made to assign participants to thermostat pairs in a way that ensured roughly even distribution of age, education, home ownership, gender, income, and confidence using a thermostat. Distributions for these variables are shown in Figure 4 through Figure 9.

Figure 4. Age distribution

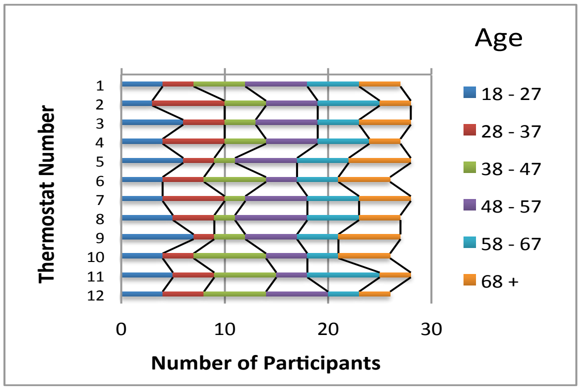


Figure 5. Education distribution

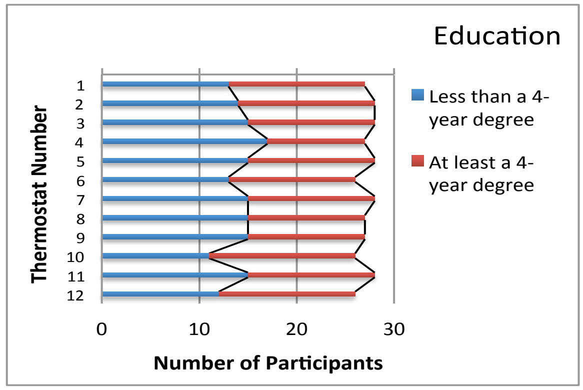


Figure 6. Income distribution

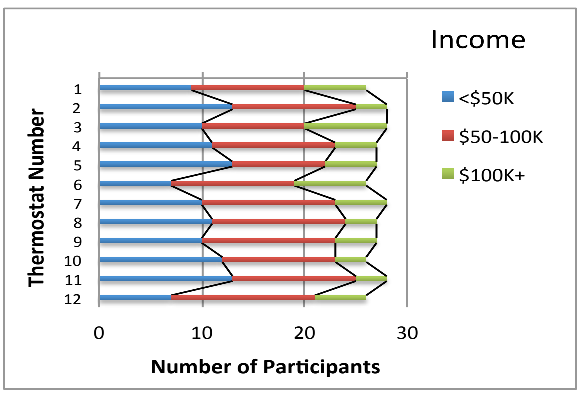


Figure 7. Home ownership distribution

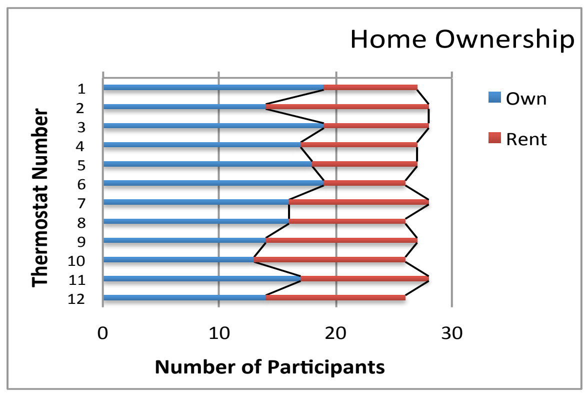


Figure 8. Gender distribution

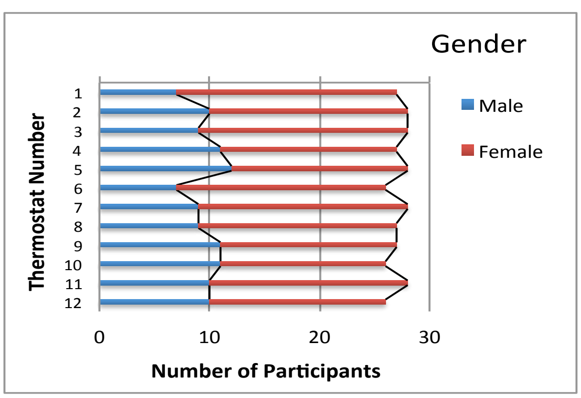
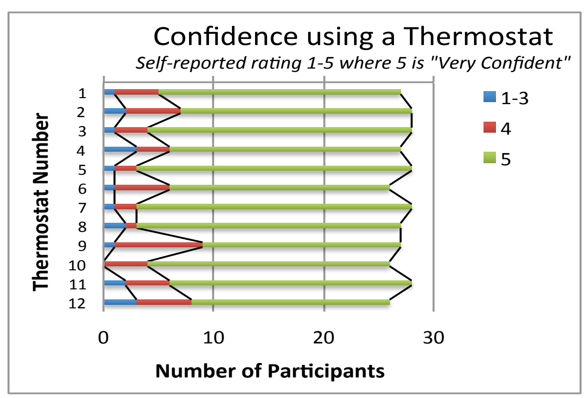


Figure 9. Confidence distribution

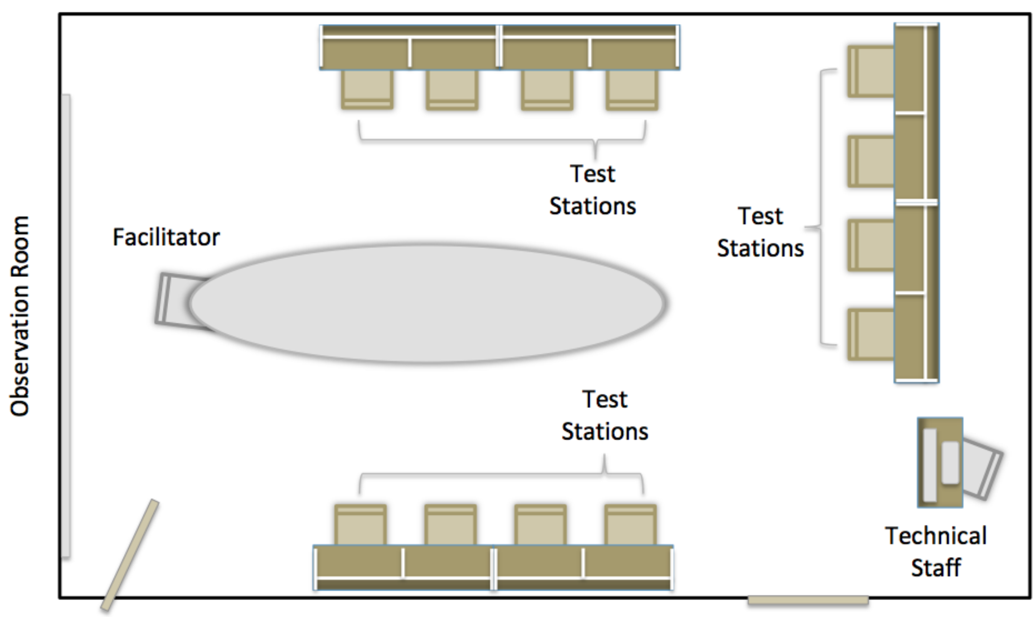


3.4 Test Lab and Equipment

The usability testing took place at a facility within the SMUD service territory. The test lab was equipped with multiple test stations affording each participant some privacy. A thermostat and its supporting applications were mounted in each cubicle, in view of the video camera. Video of participant faces was not recorded.

During testing, the facilitator and one technical support staff were seated in the same room as the participants, while observers monitored the sessions in the observation room.

Figure 10. Test room layout



Video Recording

Small cameras were affixed to each cubicle to record user interactions with the thermostats and supporting applications. One video was recorded for each of the 326 tests for later review.

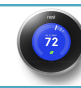
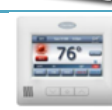
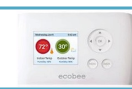
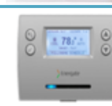
Web Access

Six of the thermostats tested required Internet connectivity – either directly or through gateways – for remote control by smartphone application or website. Both wired (Ethernet) and wireless (WiFi) Internet access was made available at the facility for use with these thermostats.

THERMOSTATS

A total of 12 thermostats were selected for testing as shown in Table 2. These units scored the highest priority ratings from a list of thermostats that are included both (a) currently used or planned for use at SMUD or (b) popular standard California Title 24 compliant thermostats readily available for sale at common retail outlets.

TABLE 2. THERMOSTAT MODELS TESTED

ID	Device	Power	Zigbee Certified	Other	Remote Access	Image
1	Lux Smart Temp	battery	(non-communicating thermostat)			
2	Honeywell FocusPro	battery	(non-communicating thermostat)			
3	RCS TZ-45 (Trane 400BB)	24V 40VA		Zwave (*)	(*)	
4	Radio Thermostat CT30 (3M-50)	16V	SEP 1.0	WiFi, usnap	Web, App	
5	Nest Learning Thermostat	24V 40VA		WiFi	Web, App	
6	Ecofactor (Computime CTW218)	24V 40VA	SEP 1.0	IP gateway	Web, App	
7	Carrier ComfortChoice Touch	24V 40VA	SEP 1.1			
8	Ecobee Smart Si	12V 1.67A	SEP 1.1	WiFi	Web, App	
9	Energate Foundation FZ100	24V 450mA	SEP 1.1	IP gateway	Web, App	
10	Energate Pioneer Z100	24V 40VA	SEP 1.1	IP gateway	Web, App	
11	Cooper-Honeywell Utility Pro	24V 40VA	SEP 1.1	Pager	(*)	
12	Emerson Smart Energy	24V 40VA	SEP 1.1	(*)	(*)	

- Third-party gateway, Web, and/or App available but not tested

3.5 Thermostat Assignment

The 163 participants each evaluated 2 thermostats for a total of 326 individual tests.

All 66 potential thermostat pairs were tested at least once, and 126 of the 132 ordered pairs were tested at least once. To avoid order bias, each thermostat was the *first* unit tested in roughly half of the tests, and the *second* unit tested in the remaining tests. Table 3 shows the final count of participants that tested each ordered pair of thermostats.

Table 3. Participant-thermostat assignments: first and second of two units tested

First Unit → Second Unit ↓	1	2	3	4	5	6	7	8	9	10	11	12	Total
1		1	1	1	1	1	1	1	1	2	1	2	**13**
2	2		1	1	1	1	1	1	1	1	2	2	**14**
3	2	2		1	1	1	1	1	1	1	2	1	**14**
4	1	2	2		1	1	1	1	1	1	1	2	**14**
5	2	2	2	1		1		1	1	1	1	1	**13**
6	1	2	2	1	2		1		1	1	1		**12**
7	1	1	2	2	2	2		1	1	1	1	1	**15**
8	1			2	2	2	2		1	1	1	1	**13**
9	1	1	1	1	2	2	2	2		1	1		**14**
10	1	1	1	1	1	1	2	2	1		1	1	**13**
11	1	1	1	1	1	1	1	2	2	2		1	**14**
12	1	1	1	1	1	1	1	2	2	1	2		**14**
Total	**14**	**14**	**14**	**13**	**15**	**14**	**13**	**14**	**13**	**13**	**14**	**12**	**163**

3.6 Procedure

Before the testing process began, participants were required to review and sign nondisclosures and recording permissions (see Appendix E). Five sessions were held each day for three days. Each session accommodated up to twelve participants. Each session took 90 minutes, roughly following the schedule shown in Table 4.

Table 4. Agenda for each session

	Segment	Minutes
1	Introduction	15
2	Thermostat test #1	20
3	Discussion #1	15
4	Thermostat test #2	20
5	Discussion #2	15
6	Wrap up	5
	Total	90

Introduction

The facilitator briefed participants on the usability test procedure, including:

- the purpose of study
- the importance of their involvement
- the facilitator's role
- the room configuration, recording systems, observers, etc.
- the testing protocol and agenda

Thermostat Testing

The facilitator briefed the participants on the testing process, stressing that the thermostats – not the participants – were being evaluated. The facilitator explained that the amount of time taken to complete each task was measured, and that exploratory behavior outside the task flow should not occur until after completion of all tasks. Participants were given 20 minutes to complete the entire task list and fill out the survey. (See Facilitator's Guide, Appendix F.)

Due to the range and extent of functionality provided in the thermostats, and the short time for which each participant was available, the tasks were designed to be the most common of available functions (Table 5, tasks 1-7). These common tasks were identical for all thermostats, with minor variations every other test to limit the need for lab staff to reset thermostats to default settings after each test. During the discussion sessions, settings were adjusted for tasks that were not successfully completed.

Where possible, each thermostat also had its own set of advanced tasks unique to that device. At the end of testing, the survey included questions designed to elicit preferences for these advanced features.

Table 5. Task list

Task	Task Booklet A	Task Booklet B
1	Identify the current indoor temperature	Identify the current indoor temperature
2	Set to cool. Identify the current target cooling temperature.	Set to cool. Identify the current target cooling temperature.
3	Change the current target cooling temperature to 79	Change the current target cooling temperature to 81
4	Identify the scheduled cooling temperature for Saturday at 8 am	Identify the scheduled cooling temperature for Saturday at 8 am
5	Set to heat. Identify the current target heating temperature.	Set to heat. Identify the current target heating temperature.
6	Change the current target heating temperature to 63	Change the current target heating temperature to 61
7	Identify the scheduled heating temperature for Saturday at 8 am	Identify the scheduled heating temperature for Saturday at 8 am
8	Advanced task 8A	Advanced task 8B
9	Advanced task 9A	Advanced task 9B

Table 6. Advanced Task 8

Thermostats	Task Booklet A	Task Booklet B
1,2,11	Set the day to Wednesday	Set the day to Saturday
3	Identify the Home energy use	Identify the Home energy use
7,12	Set the date to 1/11/13	Set the date to 2/12/12
4,5,6,8,9,10	Use the smartphone app to increase the target heating temperature by 3 degrees	Use the smartphone app to decrease the target heating temperature by 3 degrees

Table 7. Advanced Task 9

Thermostats	Task Booklet A	Task Booklet B
1,2,3,7,11,12	Set the time to 10:32 am	Set the time to 3:49 pm
4,5,6,8,9,10	Use the smartphone app to postpone heating until you get home	Use the smartphone app to postpone heating until you get home

The test commenced when users were told by the facilitator to begin. Participants were then directed to flip to the first card and begin task 1. The task ended when the participant marked that they did or did not complete the task on their task checklist. The next task began when they flip over the next task card, and so on. Participants were directed to contact facilitation staff immediately should any of the equipment fail to operate during testing.

Thermostat Surveys

Prior to beginning the test, each participant was provided a survey with their participant ID, pictures of their test thermostat, and survey questions regarding their experience with that particular thermostat. When they completed all tasks, participants were directed to complete this survey. After the second thermostat test and survey, participants were given a third survey with questions about their preference for one of the two tested thermostats and the usefulness of advanced features.

The details of the thermostat's ease of use, look, feel and sound (Table 8) were rated on 10-point Likert scales. Other questions included whether the user would recommend the thermostat to a friend or neighbor, and the perceived value of the thermostat. An example of the full thermostat questionnaire is provided in Appendix G.

Table 8. Thermostat satisfaction questions

1	Rate EASE OF USE and UNDERSTANDING
a	Information on the screen
b	Buttons, dials and switches
c	Meanings of words and symbols
d	Menu navigation
e	Overall ease of use
2	Rate how the thermostat FEELS and SOUNDS
a	Buttons
b	Touchscreen
c	Dials
d	Switches
e	Overall feel and sound
3	Rate how the thermostat LOOKS
a	Layout of the screen and buttons
b	Size of the screen
c	Color(s)
d	Readability of the smallest text
e	Overall appearance of the thermostat

After the second thermostat test, an additional set of questions was provided to ascertain (1) which of the two thermostats was preferred, and (2) ratings for the desirability of a list of potential advanced features (Table 9).

TABLE 9. SURVEY QUESTIONS TO RATE PERCEIVED USEFULNESS OF ADVANCED FEATURES

7	Do you think you would find the following features useful on a thermostat in your home?
A	**Auto-Schedule**: The thermostat programs your temperature preferences for you, based on your adjustments in the first week or two.
B	**Auto-Away**: The thermostat automatically adjusts the temperature when it senses your home is unoccupied.
C	**HVAC Energy Display**: The thermostat displays the amount of electricity used by your central heating and cooling system.
D	**Home Energy Display**: The thermostat displays the amount of energy used by your home.
E	**Efficiency Indicator**: The thermostat indicates when you adjust it to an energy efficient temperature setting.
F	**Time to Temperature**: The thermostat displays how long it will take to reach the target temperature.
G	**Online Account**: You can use a computer to adjust your thermostat settings remotely
h	**Smart phone app**: You can use a smart phone to adjust your thermostat settings remotely
i	**Color display**: The main display has more than 2 colors.
J	**Touchscreen**: The main screen is also an input device.
K	**Outdoor temperature**: The thermostat can display the outdoor temperature
l	**Price response**: The thermostat automatically adjusts settings based on your input and the price of electricity
m	**Precool**: The thermostat automatically cools your home before a high-priced peak period
n	**Proximity**: Your thermostat knows your location and automatically switches between home and away settings
o	**Parental Controls**: The thermostat allows changes to settings only after a password is provided

Group Discussions

At the end of each thermostat test and survey, a short focus group discussion took place. Lab staff used this time to check each thermostat for correct task completion and to reset the thermostats to the default initial test state for the next session as needed.

Conducting a group discussion between the two thermostat evaluations ran the risk of providing some information to participants about the thermostat tested in round two. Despite this possibility, we chose to include the discussion between the two tests for the following reasons:

- With or without the discussion between the two tests, there would be increased familiarity with the process and tasks in the second test.
- Each thermostat was tested first and second an equal number of times, so the bias inherent in the second test was evenly distributed across thermostats.
- Richer, more relevant feedback was made possible by having two discussions; i.e. the experience with the first unit may have been lost if the two tests were contiguous.
- Lab staff needed that time for logistic purposes – checking to see that the thermostats were ready for the second test.

4 Data, Analysis, and Results

For this comparison study, qualitative data, satisfaction ratings, task efficiency metrics, and preference metrics were collected for use in comparing between products and features.

4.1 Pros and Cons of tested Thermostats

Following is a list of pros and cons for each thermostat pulled from the verbatim comments left on the post-test thermostat surveys. Note that these comments, in aggregate, form the basis for the variables considered in the regression modeling described later in this paper.

Table 10. Summary of participant comments for each thermostat

ID	Thermostat	Pros (from survey comments)	Cons (from survey comments)
1	Lux Smart Temp	· Easy to use and program	· Small screen and print
		· Good labels	· Small, loud dial and switches
		· Basic instructions provided	· Dim screen
		· Good button feel	· Looks old
2	Honeywell FocusPro	· Easy to navigate	· Confusing schedule
		· Button feel and labels	· Bad brightness and contrast
		· Screen layout and size	· Buttons stick
		· Nice, simple look	· No confirmation of input
3	RCS TZ-45 (Trane 400BB)	· Easy to use	· Loud buttons
		· Easy to read, good font size	· Size and shape
		· Blue backlight is nice	· Looks old/plain
		· Good button layout	· Too many different screens
		· Energy use data	
4	Radio Thermostat CT30 (3M-50)	· Large touchscreen	· Hard to navigate
		· Backlight is nice	· Backlight goes off too quickly
		· Clear print	· Too much info on the screen
		· Smartphone App is good	· Confusing icons/symbols
			· Touchscreen not sensitive
5	Nest Learning Thermostat	· Smartphone App is good	· Screen too small
		· Modern, hi-tech, advanced	· Confusing menu
		· Looks and feels good	· Hard to get started
		· Dial is nice, simple to use	
		· Would be easy to use with practice	
6	Ecofactor-Computime CTW218	· Backlight is nice	· Button feel and loudness
		· Can use with smartphone, PC	· Button sizes, shapes, layout
			· Size of the thermostat
			· Inconsistent interfaces for app, thermostat, and computer

(continued from the previous page)

ID	Thermostat	Pros (from survey comments)	Cons (from survey comments)
7	Carrier Comfort Choice Touch	· Touchscreen and Colors · Easy to use and program · Large, well lit, easy to read · Button feel and sound · Looks nice	· Button layout · Thermostat too big
8	Ecobee Smart Si	· Sleek, colorful, modern · Bright, easy to read · Button feel and sound · Home button · Easy to use and navigate	· Small display · App is difficult to use · Confusing, cluttered menu · No Help button · Symbols and labels need explaining
9	Energate Foundation FZ100	· Button size, shape, and feel · Energy saver and pricing · Easy to navigate · Easy to read · Information on the screen	· Too many buttons · Small screen · Multi-function of some buttons · Boring colors · Not modern looking
10	Energate Pioneer Z100	· Easy to navigate · Button feel and sound · Backlight · Smartphone App	· Button layout · Hard to navigate menu · Multi-function of some buttons · Small print · Not bright enough
11	Cooper-Honeywell Utility Pro	· Touchscreen · Screen color and backlight · Size · Information on the display	· Hard to read screen · Slow response · Buttons too small · Small text · Poor contrast
12	Emerson Smart Energy	· Large font, screen, easy to read · Easy to use and navigate · Bright, blue backlight · Button layout · Overall appearance	· Hold is confusing · Button size and feel

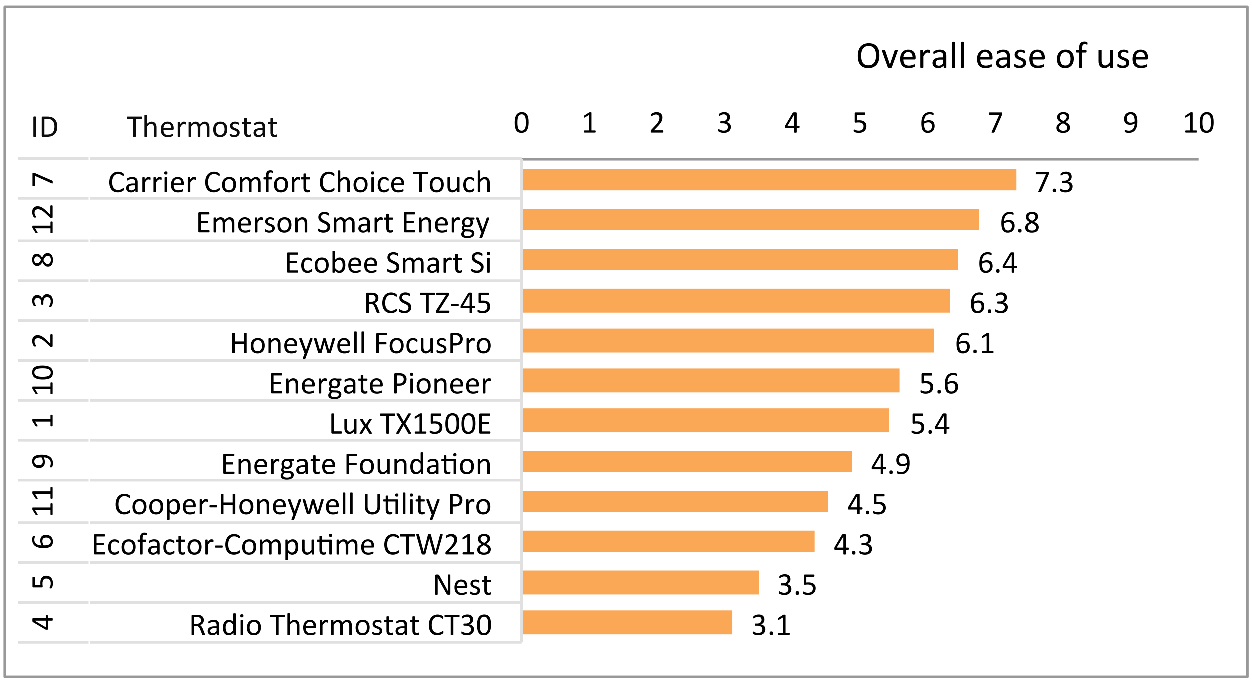
4.2 Feature Ratings for tested Thermostats

After testing each thermostat, participants filled out surveys as described in Section 3.6. This section provides the average thermostat ratings for the overall ease of use, feel and sound, and appearance as rated by participants in these surveys.

Ease of Use

Figure 11 ranks the 12 thermostats tested for this study by participant scores for "Overall Ease of Use" (Table 8, question 1e). Of the twelve, the Carrier ComfortChoice Touch garnered the highest average rating, statistically outperforming the bottom four rated thermostats: the Radio Thermostat CT30, Nest, Ecofactor/Computime CTW218, and Cooper/Honeywell Utility Pro. Other thermostats that were highly rated for ease of use include the Emerson Smart Energy, the Ecobee Smart Si, the RCS TZ-45, and the Honeywell FocusPro.

Figure 11. Usability ratings for all thermostats tested



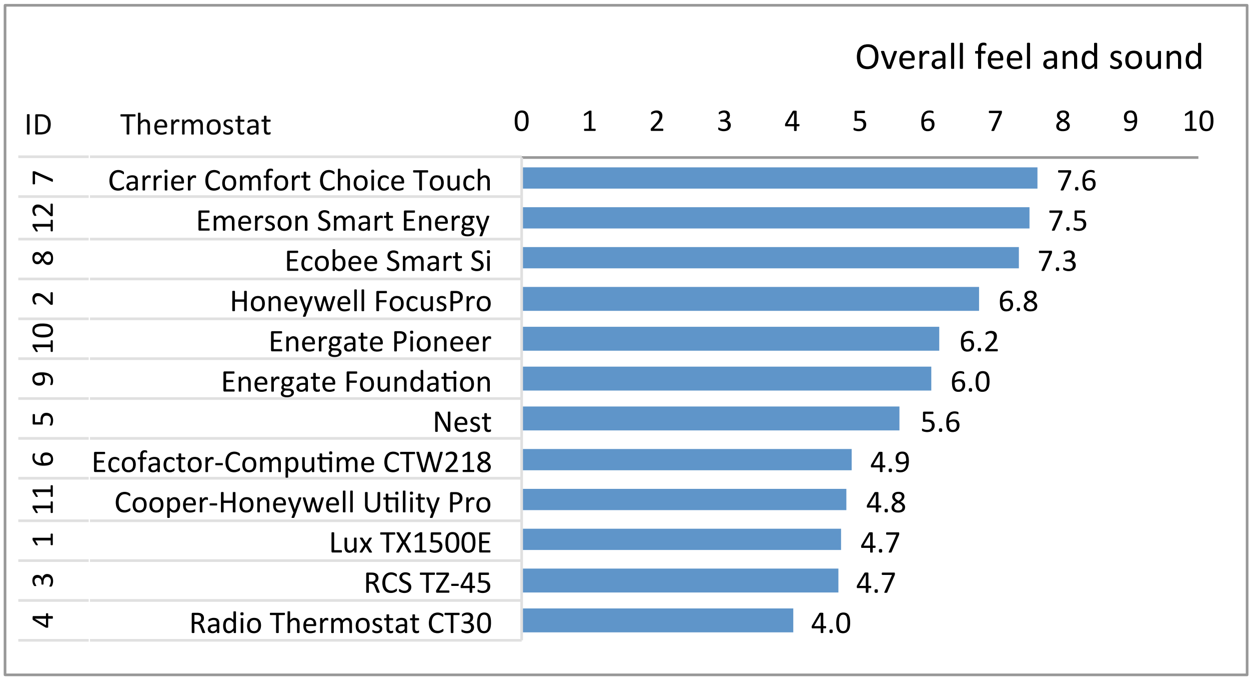
Statistical significance bounds: ± 2.5 (α=0.01)

The Energate Pioneer, which is being used in two SMUD pilots in summer 2013, ranked sixth in this category. The Ecofactor and Nest thermostats, also being used in 2013 pilots, ranked a disappointing 10th and 11th out of 12 in the Ease of Use category. For a discussion of the low Ease of Use scores earned by the Nest Learning Thermostat, see section 5.3.

Feel and Sound

Figure 12 shows that the Carrier ComfortChoice Touch also took first place in "Feel and Sound" (Table 8, question 2e) – the category that turns out to be the most significant in predicting participant preferences, as will be seen in the following pages. In this category, the Carrier statistically outperformed the bottom five rated thermostats: the Radio Thermostat, RCS TZ-45, Lux TX1500E, Cooper/Honewell Utility Pro, and Ecofactor/Computime CTW218. Other thermostats that were highly rated for feel and sound include the Emerson Smart Energy, the Ecobee Smart Si, and the Honeywell FocusPro.

Figure 12. Feel and Sound ratings for all thermostats tested



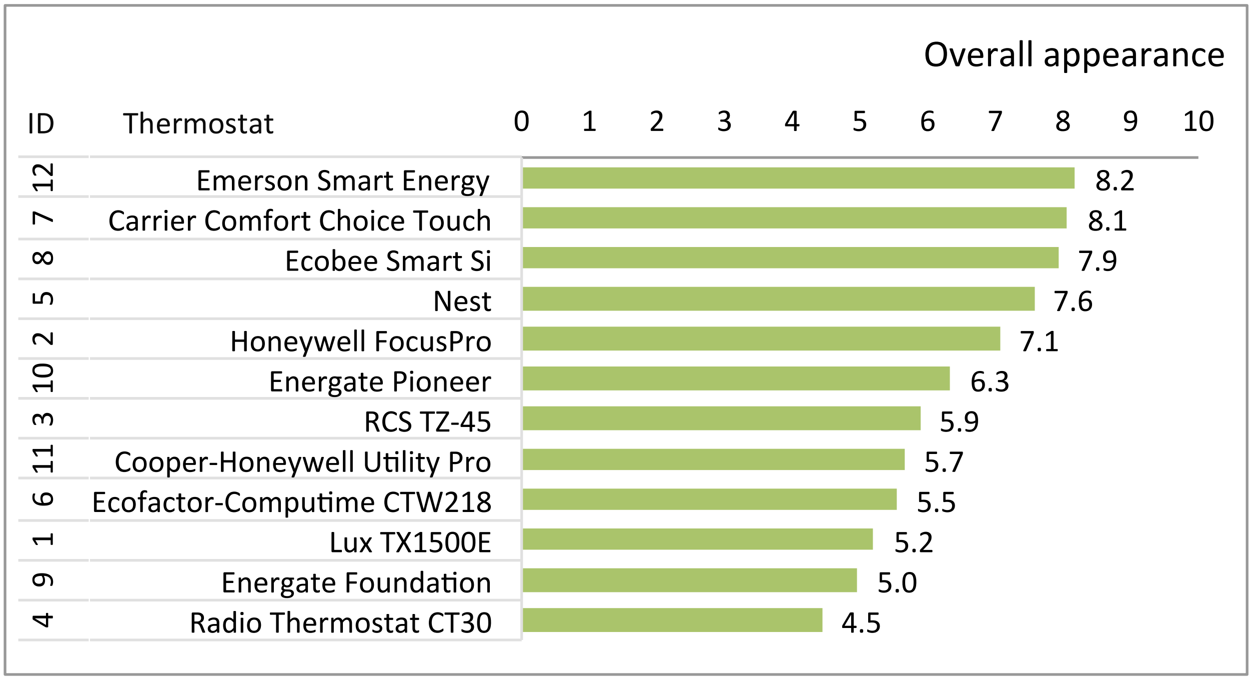
Statistical significance bounds: ± 2.6 (α=0.01)

The Energate Pioneer, which is being used in two SMUD pilots in summer 2013, ranked fifth in this category. The Nest and Ecofactor thermostats, also being used in 2013 pilots, ranked 7th and 8th out of 12 in the Feel and Sound category.

Appearance

Figure 13 ranks the thermostats tested for this study by participant scores for "Overall Appearance" (Table 8, question 3e). The Emerson Smart Energy took first place, statistically outperforming the bottom five rated thermostats: the Radio Thermostat, Energate, Lux, Ecofactor, and Cooper. Other thermostats highly rated for Appearance include the Carrier ComfortChoice Touch, the Ecobee Smart Si, the Nest Learning Thermostat, and the Honeywell FocusPro.

Figure 13. Appearance ratings for all thermostats tested



Statistical significance bounds: ± 2.4 (α=0.01)

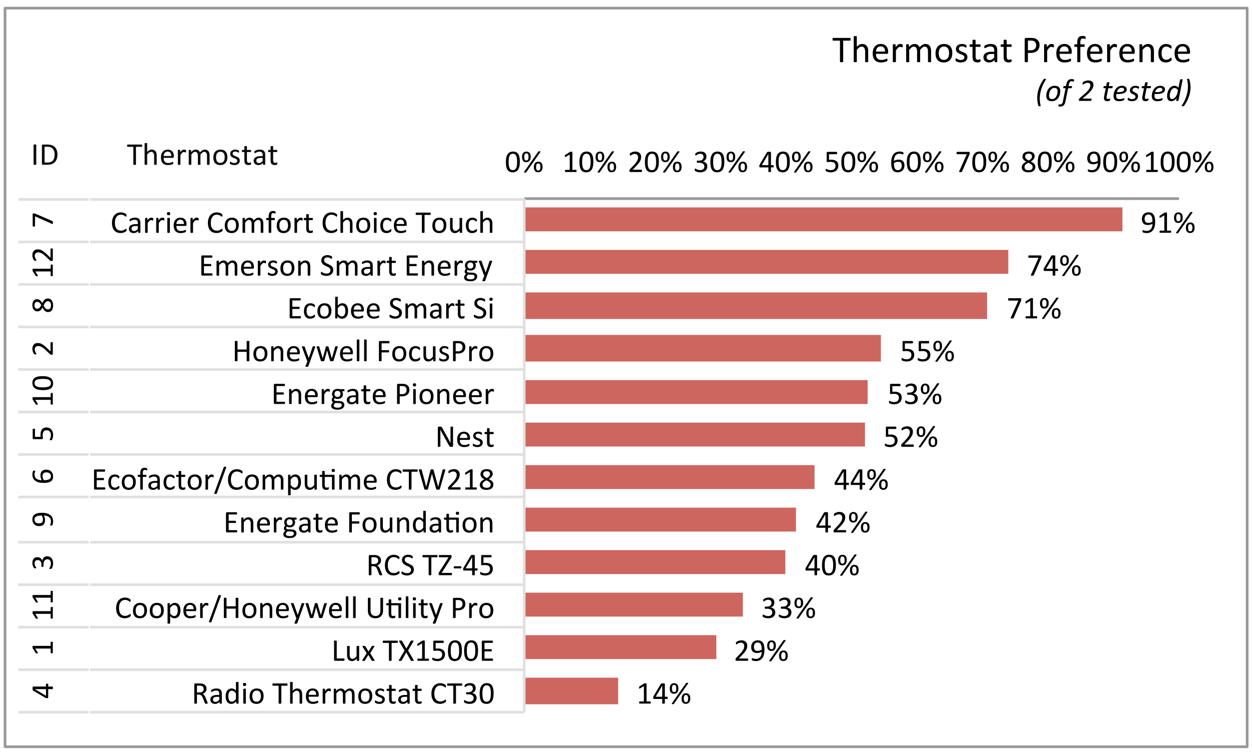
Of the thermostats used in SMUD field pilots in 2013, the Nest thermostat ranked 4th, the Energate Pioneer ranked 6th, and the Ecofactor ranked 9th out of 12 in the Appearance category.

4.3 Preferences for Tested Thermostats

In the survey provided after the second thermostat test, participants were asked to choose the thermostat they would purchase given the choice of the two units they had tested. More than 90% of participants that tested the Carrier ComfortChoice Touch chose it as their preferred thermostat (Figure 14). Of the 22 participants that chose the Carrier, 18 participants cited among their reasons: the ease of use (11 participants), the touchscreen (5 participants), the appearance (4 participants), the clarity and size of the font (4 participants), the color display (2 participants), and advanced features (2 participants).

Also very popular were the Emerson Smart Energy and the Ecobee Smart Si, while the Radio Thermosats CT30, Lux TX1500E, and Cooper/Honeywell Utility Pro ranked in the bottom three. The Energate Pioneer, which is being used in two SMUD pilots in summer 2013, was the fifth most preferred thermostat of the twelve.

Figure 14. Preference scores for all thermostats tested



Statistical significance bounds: ± 60% (α=0.01)

4.4 Task Efficiency scores for tested thermostats

All 326 video recordings of the usability tests were reviewed to capture Time-on-task and Success measures for each thermostat, participant, and completed task. These measures were then used to calculate the individual and average Task Efficiency metric for each thermostat.

For all tasks marked "Done" on the task checklist, the task start time was recorded at the moment the numbered task card became visible by the camera, and the end time was recorded as the participant marked their checklist. In a few cases, participants did not complete the task on the first try, marked their checklist Not Done, then returned to and completed the same task later in the session. In these cases, the two times were added together to obtain the total Time-on-task metric.

Successful completion of the task was also determined using the video recording. Tasks successfully completed received a Success score of 1, while those that were not successfully completed received a Success score of zero, even if the participant marked that task on their checklist "Done."

Together, the Time-on-task and Success metrics were used to calculate the *Task Efficiency* metric, defined on a scale from 0% to 100%, such that 0% indicates that the task could not be completed at all, and 100% indicates successful completion in no time (Eq. 1). A similar metric is described in Perry et al., 2011.

$$\text{Task Efficiency} = 2s / (1+e^{t}) \qquad (1)$$

Where

- s = Success = {0 for failed tasks; 1 for completed tasks)
- t = Time-on-Task = time to complete the task, in minutes.

Figure 15. Task Efficiency metric ranges from 100% to 0%

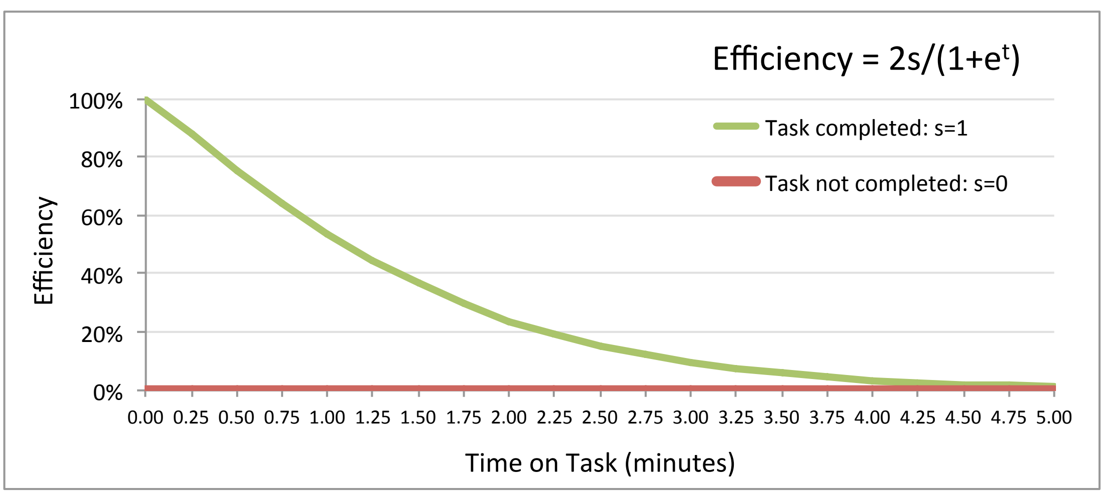
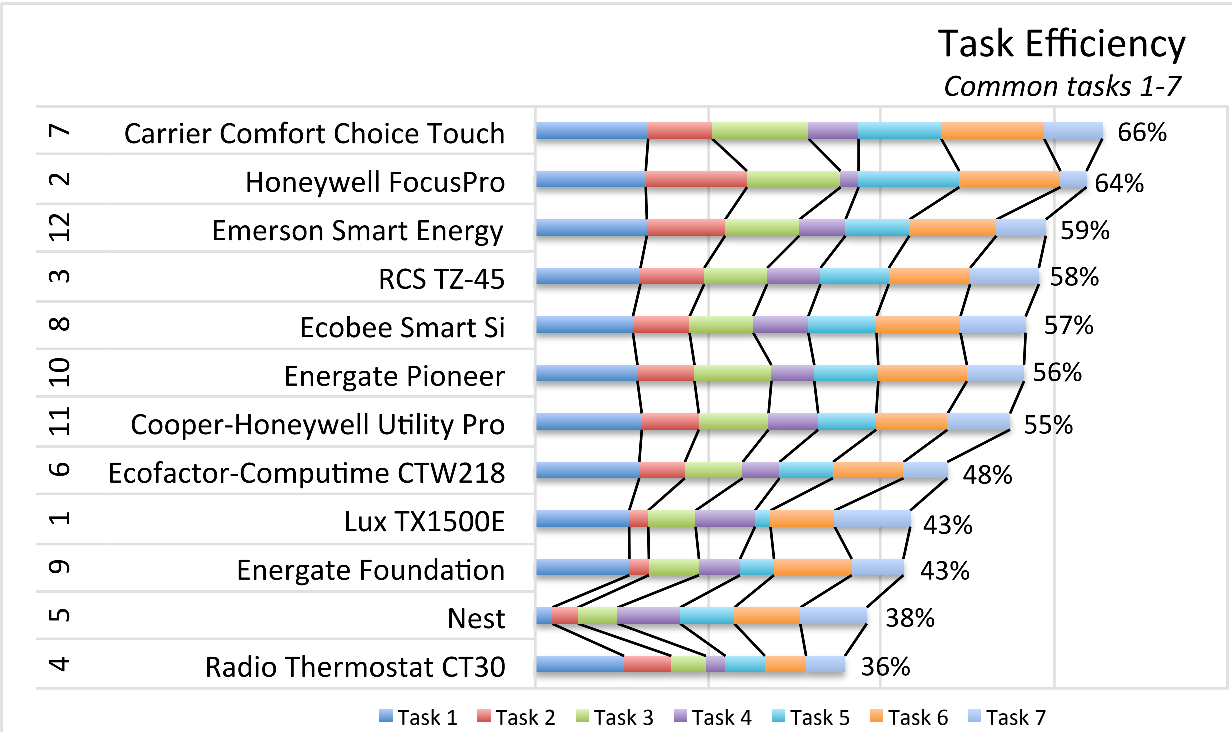


Figure 16 ranks the twelve thermostats tested in this study by final efficiency scores, showing proportional contributions from each task. The top ranked thermostat is the Carrier ComfortChoice Touch, followed closely by the Honeywell FocusPro. Scoring below 50% were the Radio Thermostat, Nest, Energate Foundation, Lux, and Ecofactor/Computime thermostats. The Energate Pioneer, currently in the field for two pilots, scored sixth.

FIGURE 16. TASK EFFICIENCY SCORES FOR ALL THERMOSTATS TESTED



Statistical significance bounds: ± 12% (α=0.01)

Visually striking in Figure 16 is the unusually low score for Task 1 of thermostat 5, which asked for the current indoor temperature on the Nest learning thermostat. This low efficiency score is mainly a result of participants mistaking the large number in the center of the Nest for the current indoor temperature, when in fact it represented the target temperature. The actual current indoor temperature value is revealed only when the face of the thermostat is pushed or turned, and then it is displayed as a much smaller number located on the dial's perimeter. Another contributing factor was the fact that participants were not provided with any instructions on how to use the thermostats, so it took a relatively long amount of time for many of them to realize that the entire thermostat face was to be pushed and turned as the input mechanism. In fact, more than half of participants that tested the Nest were unable to come to this realization during the 20-minute test period. This issue is further discussed in section 5.3.

4.5 Regression Models for Preference and Efficiency

To investigate the effects of thermostat features and participant characteristics on participant thermostat Preference and task Efficiency, two ordinary linear regression models were implemented. The first model regressed the Preference scores shown in Figure 14 on the fifteen variables shown in Table 11, selected to represent the full set of features and participant characteristics while avoiding multicollinearity. The second model regressed the Efficiency scores shown in Figure 16 on the same fifteen variables (Table 11).

Table 11. Regression model variables

Variable Type	Description	Data type
Participant Characteristics	Household income less than $50,000	Boolean
	Household income $50,000-$100,000	Boolean
	Age (18+)	Continuous
	Education: more or less than a 4-year college degree	Boolean
	Gender	Boolean
	Homeowner or renter	Boolean
	Self-rated confidence using a smartphone	Continuous
	Self-rated confidence using a thermostat	Continuous
Thermostat Features	Remote control via SmartPhone App	Boolean
	Overall appearance rating (survey question 3e)	Continuous
	Overall ease of use rating (survey question 1e)	Continuous
	Overall feel and sound rating (survey question 2e)	Continuous
	Color display screen (more than 2 colors)	Boolean
	Screen size in square inches	Continuous
	Touchscreen	Boolean

Preference Model Results

Based on the Preference model output, none of the 8 participant characteristics shown in Table 11 had a significant impact on thermostat preference. More important in participant preferences were the thermostat features. Of the 7 thermostat features included in the model, the following two significantly increased the likelihood that a participant would choose one thermostat over another:

1. Good overall feel and sound (p=0.002)
2. Color displays (p=0.008)

For the full Preference model output, see Appendix H.

Efficiency Model Results

Based on the Efficiency model output, 2 of the 7 thermostat features from Table 11 were associated with significantly higher task efficiency scores:

1. Higher ratings for ease of use (p<0.0001)
2. Larger screens (p=0.002)

Efficiency was also influenced by 2 of the 8 participant characteristics from Table 11. In particular, this study showed significantly lower efficiency scores for:

1. Older participants (p<0.0001)
2. Renters (p=0.003)

Notably absent from this list are variables for income, education, and gender. In fact, of the 15 variables included in the regression model, these had the lowest effect on Task Efficiency, having p-values greater than 0.60 in all cases. Other factors that were found to be unrelated to Task Efficiency scores were self-rated confidence using thermostats and smart phones.

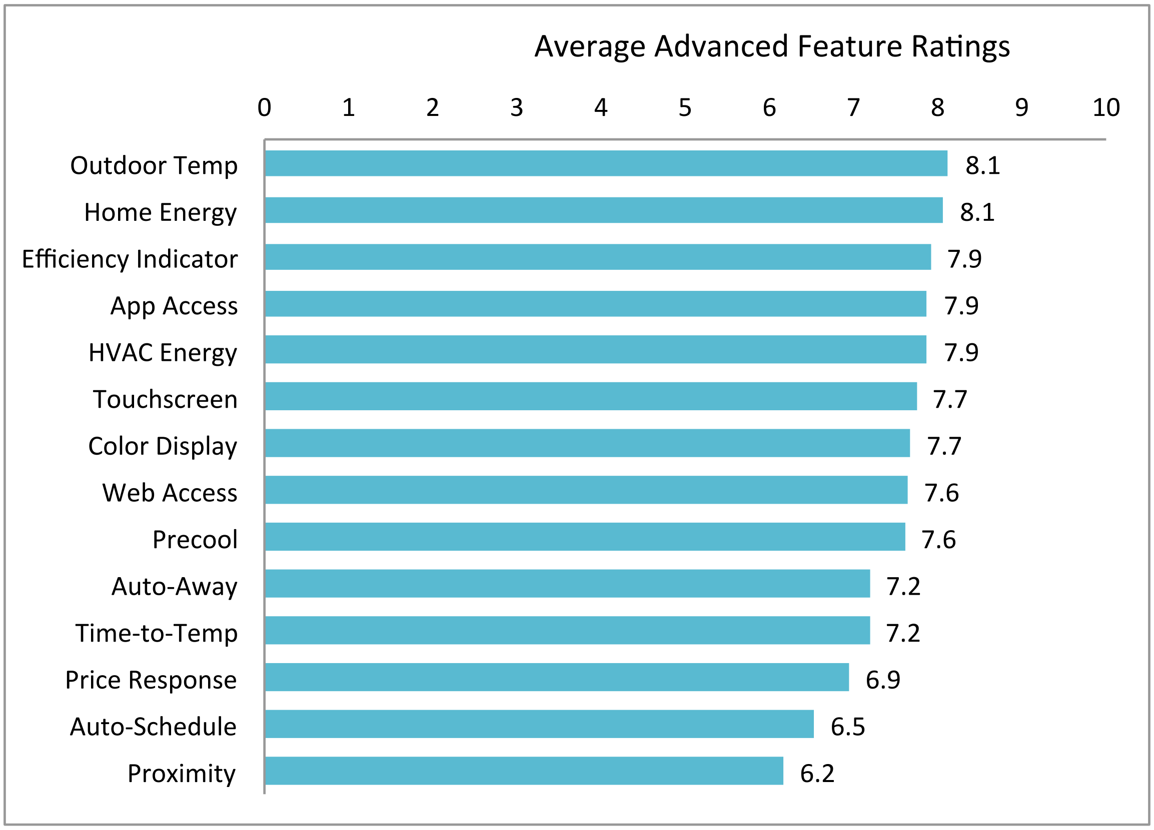
For the full Efficiency model output, see Appendix H.

4.6 Advanced Feature Ratings

The second survey, completed after the completion of both thermostat tests, presented participants with the question: "Do you think you would find the following features useful on a thermostat in your home?" Possible responses ranged from 1, Not at all, to 10, Definitely. (See Table 9 for the full list of advanced features, or Appendix G for the full survey.)

Overall, the ability to see Outdoor temperature and real-time Home Energy use on the thermostat garnered the highest ratings – however, efficiency indicators, smartphone access, HVAC energy data, touchscreens, color displays, web access, precooling functionality, auto-away, time-to-temp, and price response functionality scored statistically similar ratings. Only Auto-schedule and proximity features scored statistically lower than the top rated features.

FIGURE 17. PERCEIVED USEFULNESS OF 15 ADVANCED FEATURES

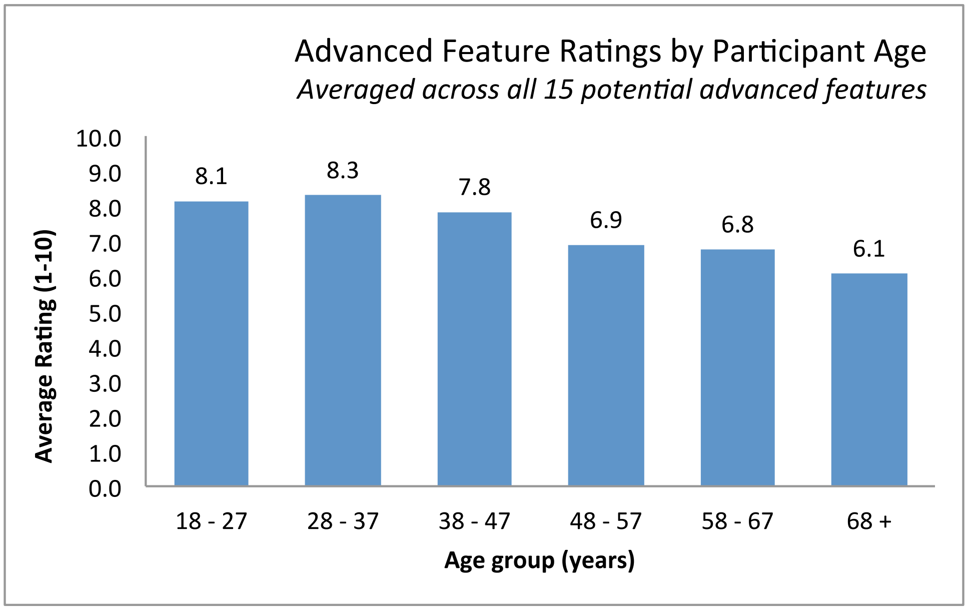


Statistical significance bounds: ± 1.3 (α=0.01)

Advanced Feature Preferences by Age

These advanced feature ratings become somewhat more interesting when considered alongside the age of the participant. In general, younger users were much more likely to consider any of the advanced features useful, as indicated by the average ratings provided in Figure 18.

Figure 18. Average ratings across all 15 advanced features, by age

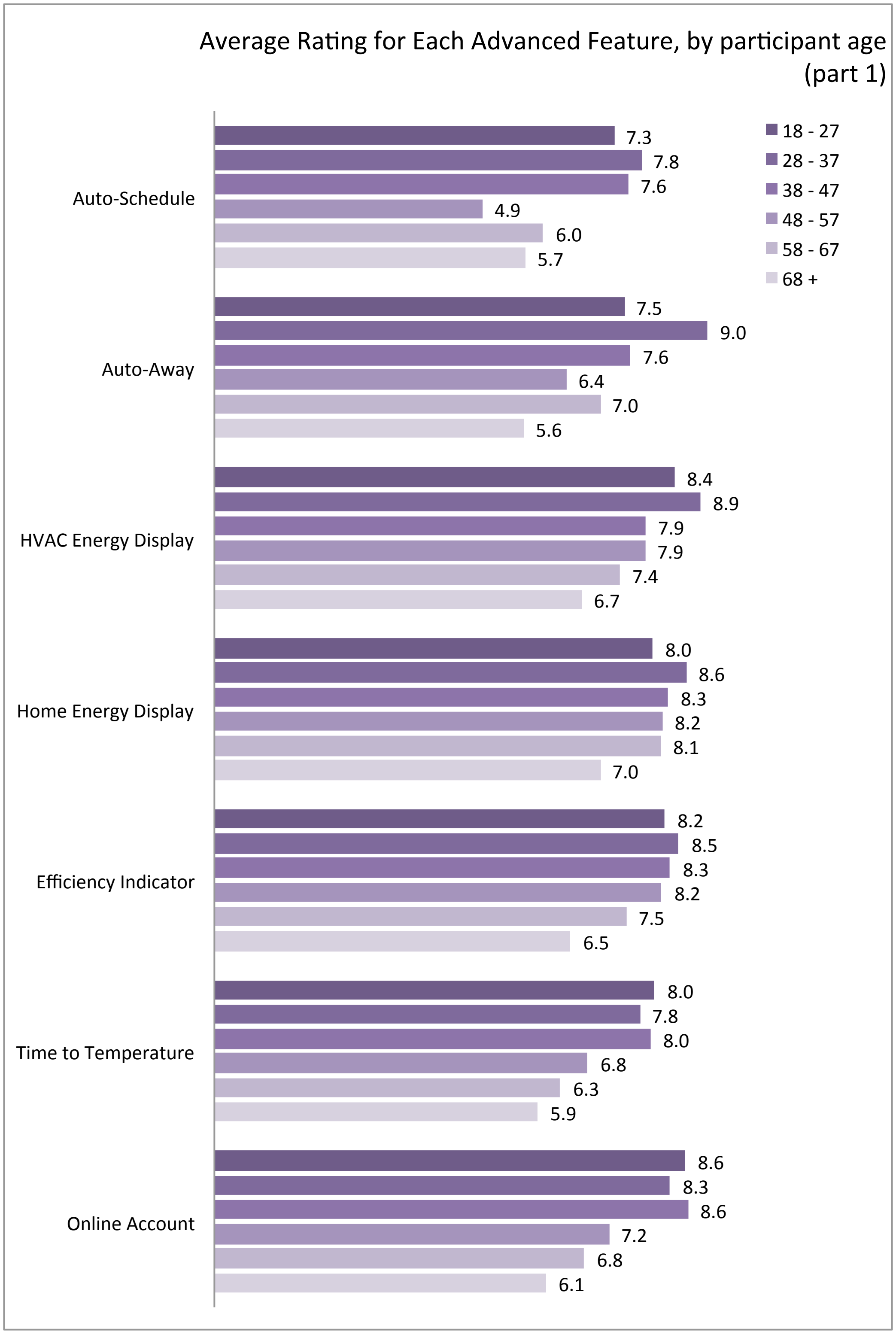


On a feature-by-feature basis (Figure 19), the ratings are surprisingly consistent, with younger participants generally giving higher ratings and older participants generally giving lower ratings.

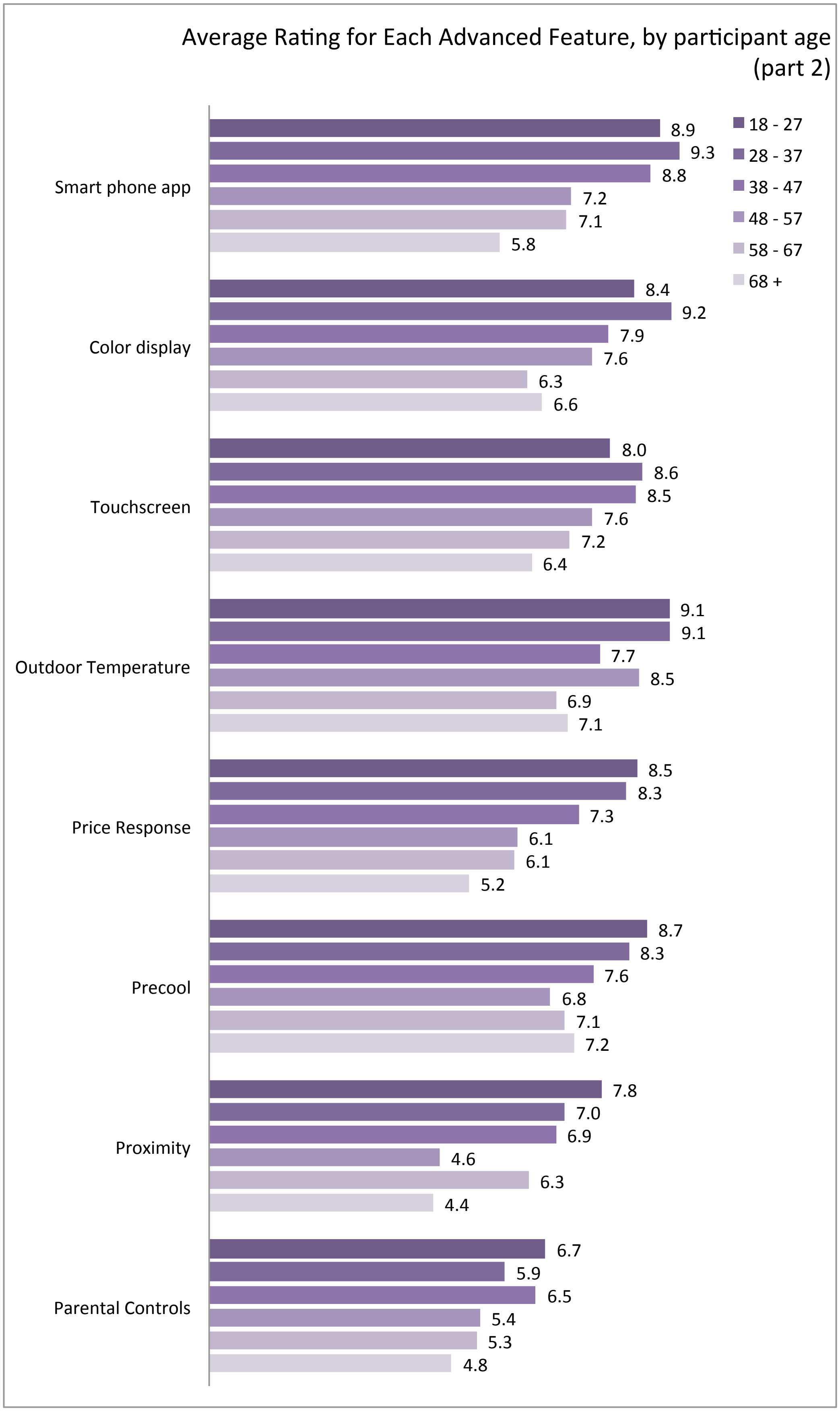
Note that the Home Energy Display not only has one of the highest average ratings, but is also the most universally appreciated, being the only advanced feature with average ratings at or above 7.0 in every age category.

At the other end of the spectrum, Auto-Away functionality and Smartphone applications had the largest spread in ratings, with the 28 to 37 years olds being the most enthusiastic, and the participants 68 years and older being the most uninterested.

FIGURE 19. AVERAGE RATING FOR EACH ADVANCED FEATURE, BY PARTICIPANT AGE



(Figure 19 is continued on the next page.)


Average Rating for Each Advanced Feature, by participant age
(part 2)
18 - 27
28 - 37
38 - 47
48 - 57
58 - 67
68 +
Smart phone app
8.9
9.3
8.8
7.2
7.1
5.8
Color display
8.4
9.2
7.9
7.6
6.3
6.6
Touchscreen
8.0
8.6
8.5
7.6
7.2
6.4
Outdoor Temperature
9.1
9.1
7.7
8.5
6.9
7.1
Price Response
8.5
8.3
7.3
6.1
6.1
5.2
Precool
8.7
8.3
7.6
6.8
7.1
7.2
Proximity
7.8
7.0
6.9
4.6
6.3
4.4
Parental Controls
6.7
5.9
6.5
5.4
5.3
4.8

Other Advanced Feature Requests

Beyond the 15 advanced features we asked participants to rate, some users used the discussion sessions to suggest a few additional thermostat features they would like. These included:

- Ability to enlarge the font or zoom in
- Ability to integrate with other devices, e.g. whole house fans
- Ability to set a budget for heating/cooling and have thermostat automatically adjust
- Customer programmable screen color options
- Decorative housing skins
- Display bill balance on thermostat
- Help videos and demos
- Motion sensing to turn backlighting on and off automatically
- Simple climate control dials, such as those commonly found in cars
- Sound options – type, tone and volume
- Voice control
- Wireless unit that can be placed anywhere in the home

5 Summary and Conclusions

The goal of this study was to link thermostat features and participant characteristics with Task Efficiency and Thermostat Preference.

Data was collected during a three-day lab study during which 163 participants, representing 6 age groups and 2 education levels, each performed 7 identical tasks on 2 of the 12 thermostats tested in this study. Each of the 12 thermostats was tested by between 26 and 28 participants, roughly evenly distributed by age, education, income, home ownership, and gender.

Surveys collected user ratings for test thermostats' ease of use, feel and sound, and appearance, along with participant preference for one of the two thermostats tested and ratings for 15 potential advanced features. Videos of individual thermostat tests were used to determine time-on-task for each of the 7 tasks.

Time-on-task values were then used to calculate an Efficiency metric for each task and thermostat on a scale from 0 to 1. Thermostat Preference was calculated as the percentage of participants that chose that thermostat from the two they tested. Average Efficiency and Preference scores were used as the dependent variables in separate linear regression models that included 15 thermostat features as independent variables.

The main findings of this study are as follows.

Preference scores were significantly higher for thermostats with:

- Good overall feel and sound (p=0.002)
- Color displays (p=0.008)

Preference scores were similar across participants of differing age, gender, education, income, home ownership, and technology IQ.

Efficiency scores were significantly higher for thermostats with:

- Higher ratings for ease of use (p<0.0001)
- Larger screens (p=0.002)

Efficiency was also influenced by the characteristics of the user. In particular, this study showed significantly lower efficiency scores for:

- Older users (p<0.0001)
- Renters (p=0.003)

The remainder of this section provides a discussion of these findings by feature, a review of the 12 thermostats from the perspective of these findings, and finally, a list of recommendations for future thermostat programs at SMUD.

5.1 Results by Feature

SmartPhone App

The existence of a smartphone app did not improve or degrade the Efficiency or Preference scores for this study. Many participants, both in the surveys and in the discussion sessions, explicitly mentioned the smartphone app as a positive feature, however, there were also those who said they would not use a smartphone app at all.

In reviewing survey data for advanced features (Figure 19), one can see that there is a large discrepancy between the younger and older participants: on average, the 28 to 37 year-old group rated the smartphone app a 9.3 out of 10 possible points, but the 68 and over group rated it just 5.8 out of 10. In fact, of all the advanced features, the smartphone app was the one with the largest age discrepancy.

Appearance

In the discussion sessions and survey comments, many participants commented on the appearance of the thermostats. Some particularly disliked thermostats that are taller than they are wide, while others were more concerned with overall size or how far it stuck out from the wall. Many said they preferred those that look more modern, and pointed out those in the room that appeared too old-fashioned for their taste. With the exception of screen size, however, appearance was not a significant predictor of Preference or Efficiency. Individual appearance variables including layout of the screen and buttons, colors, and readability of the smallest text were also not significant in the models.

Number of physical buttons

The number of buttons was not a significant predictor of Efficiency or Preference. In the discussion session and survey comments, comments on the number of buttons were very few, suggesting that this feature is not particularly important to customers. As further evidence, one could look at the thermostats studied to find that the two most popular thermostats had very different numbers of buttons. The most Preferred thermostat, the Carrier ComfortChoice Touch, had just 3 buttons, while the second most preferred thermostat, the Emerson Smart Energy, had the most of any of the units at 10. A likely confounding factor in this variable is that touchscreen “buttons” could not be counted as buttons because the number of these changed depending on the available menu options. Because of these confounding factors, this variable was omitted in the final model.

Feel and Sound

While ratings for the individual "feel and sound" of buttons, dials, switches and touchscreens were all collected in the thermostat surveys, it was the rating for the "overall feel and sound" of the thermostat that was by far the strongest predictor of Preference. In fact, these individual ratings were ultimately dropped from the models due to missing data for thermostats without these input devices and multicollinearity with the "overall feel and sound" rating.

Participant Characteristics

During the recruitment process, participants were asked to provide information about their age, education level, income, and confidence using a variety of technologies. Of these, only those variables describing age and home ownership were associated with significant impacts.

Age

Participants in this study were recruited to fill six ten-year age groups: 18-27, 28-37, 38-47, 48-57, 58-67, and 68+. Regression analysis indicated that higher ages were strongly associated with lower Task Efficiency scores, but had no influence on thermostat Preference. Age did, however, influence preferences for specific advanced features, with older participants generally being much less interested in advanced features than their younger counterparts (see Figure 18, Figure 19).

Home ownership

Home ownership was a significant predictor for Task Efficiency, with renters scoring significantly lower than homeowners. Since efficiency was not related to income or education levels, a possible explanation for the effect of home ownership is that renters have been exposed to older, cheaper, and less user-friendly thermostats than have homeowners. One might further theorize that they have had few positive experiences navigating thermostats, and in fact, may be more likely to use them as on-off switches rather than attempting to master a new device each time they move.

Gender, Education, and Income

Variables for gender, education and household income were not significant in the regression model.

Confidence with Technology

Participants' self-reported confidence with smartphones and thermostats had no significant impact on Task Efficiency or Thermostat Preference.

Screen

Color display

Color display was associated with significantly higher scores for Thermostat Preference. This feature was also frequently mentioned in the survey comments and discussion sessions as being a preferred feature. In the advanced feature ratings, color display ranked seventh, but was in a statistical dead-heat with the most popular of the advanced features.

Touchscreen

Somewhat surprisingly, thermostats with touchscreens scored significantly lower on Task Efficiency and Thermostat Preference, despite a fairly high rating for touchscreens in the advanced features survey. Based on participant comments, this outcome is likely due to the touchscreen thermostats without color displays – the Radio Thermostat and the Cooper/Honeywell Utility Pro, both of which were in the bottom three preferred thermostats studied.

Screen size in square inches

A larger screen size was associated with significantly improved Task Efficiency scores – but somewhat unexpectedly, not with improved Preference scores. Screen sizes of the 12 thermostats included in this study ranged from 2.4 square inches for the highly preferred Ecobee Smart Si with a Preference score of 71%, to 10.0 square inches for the Cooper/Honeywell Utility Pro, which had Preference score of just 33%. As indicated by the results of the regression analysis, the color display and feel and sound of the Smart Si were more important to user Preference than was the larger screen size of the Cooper-Honeywell Utility Pro.

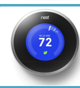
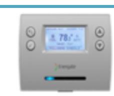
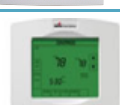
Ease of Use

Although the ease of use variables were highly correlated with both Efficiency and Preference metrics, this significance disappeared once other variables were included in the model. Thus, all else being equal, the information on the screen, usability of buttons, dials and switches, meanings of words & symbols, and menu navigation would be important factors. However, given the real-world variety of screen type and size, and the physical quality of thermostats interpreted through feel and sound, these ease of use variables are not significant predictors of Efficiency or Preference.

5.2 Results by Thermostat

Table 12 lists the pros and cons for each of the 12 thermostats, focusing only on the features that contribute significantly to the Efficiency and/or Preference for each thermostat. For a summary of customer comments without regard to statistical significance, see Table 10.

Table 12. Summary of statistically significant pros and cons of tested thermostats

ID	Device	Image	Preference Rank	Efficiency Rank	Pros	Cons
7	Carrier Comfort Choice Touch		1	1	Large screen Good feel/sound Color display Easy to use	
12	Emerson Smart Energy		2	3	Good feel/sound Large screen Easy to use	Mono display
8	Ecobee Smart Si		3	5	Good feel/sound Color display Easy to use	Small screen
2	Honeywell FocusPro		4	2	Good feel/sound Large screen	Mono display
10	Energate Pioneer Z100		5	6	Good feel/sound	Mono display
5	Nest Learning Thermostat *		6	11	Good feel/sound Color display	Small screen Not easy to use
6	Ecofactor-Computime CTW218		7	8		Bad feel/sound Mono display Not easy to use
9	Energate Foundation FZ100		8	10	Good feel/sound	Small screen Mono display
3	RCS TZ-45 (Trane 400BB)		9	4	Easy to use	Bad feel/sound Mono display
11	Cooper-Honeywell Utility Pro		10	7	Large screen	Bad feel/sound Mono display Not easy to use
1	Lux Smart Temp		11	9		Bad feel/sound Small screen Mono display
4	Radio Thermostat CT30 (3M-50)		12	12		Bad feel/sound Mono display Not easy to use

** See the next page for a discussion about the Nest Learning Thermostat.*

5.3 Design concerns for SMUD thermostat pilots

Of the twelve thermostats tested for this study, three are currently in use in SMUD field pilots: The Nest Learning Thermostat, the Ecofactor-Computime CTW218, and the Energate Pioneer. Following are design considerations specific to these three thermostats intended to help SMUD project managers develop better customer education materials related to these units.

Nest Learning Thermostat

During testing, the Nest Learning Thermostat was the only unit that generated unsolicited attention and interest. Many participants were intrigued by its sleek, round, retro look and blue glow – so much so that some even protested that they did not get the opportunity to test it.

Those who tested the Nest, however, were ambivalent. While the cold glass, metallic disc, and iPhone like clicks scored highly on the "feel and sound" category, more than half of the participants that tested the Nest – 16 of 28 – were unable to figure out the input mechanism at all or until the very end. Because of this, the Nest garnered a very low 38% Task Efficiency score.

Excluding the 16 participants that were baffled by the push and turn input mechanism, the Efficiency score for the Nest would have been a very respectable 68%, placing this thermostat in the number one rank for Efficiency. As measured for this study, the Nest ended up in the number 11 position for Efficiency, because most participants that tested the Nest were unable to complete more than one or two tasks. Excluding these same participants from the Preference analysis would have given the Nest a 67% Preference score, moving it into the number 4 position after the Carrier, Emerson, and Ecobee.

Also an issue for the Nest is the placement and size of the target temperature relative to the current indoor temperature. Whereas the current indoor temperature is generally the largest temperature value displayed, the Nest shows the target temperature in large font in the center of the thermostat. As a result, the vast majority of Nest testers entered the target temperature when the current indoor temperature was requested.

Negative comments collected in the surveys focused almost exclusively on the difficultly figuring out the input mechanism and the desire for a user manual, although several mentioned the small screen size.

ECOFACTOR/COMPUTIME

The main complaint of participants that tested this unit had to do with the button size, placement and functionality. In particular, the large silver up and down buttons on the right side of the unit were considered unattractive by many, while a few did not recognize them as buttons at all until well into the timed testing period. One participant complained that three different button sizes and shapes was too many.

As or more problematic was the fact that the large buttons imprinted with symbols for up and down did not increase and decrease the target temperature, as many participants had expected. Instead, the small round button at the bottom of the unit changed the target temperature with a left and right pressing motion. As a result, many participants had trouble determining, in one participants words, "whether to use the up down arrow or the side to side arrow," which in turn made the navigation of menus "puzzling."

Another very common complaint was that the thermostat interface was not consistent with the smartphone app or the computer interface. In essence, this required that the participant learn multiple interfaces for the same appliance, which in some cases was "totally confusing."

One of the customers inadvertently pressed the up and down buttons at the same time, initiating a lockout function, and then was unable to complete any further tasks. Incidentally, this also happened to the technical staff during setup, requiring a call to customer service for resolution.

Not mentioned in the comments, but seen in the video, was an issue in the navigation that required participants to press the "Select" button before moving on to the next menu. To unselect that choice, they had to hit "Select" again. This caused a reasonable amount of confusion during menu navigation for many participants.

Energate Pioneer

The most common complaints of the Energate Pioneer were related to the layout and functionality of the buttons. There were also several comments indicating confusing menus and dissatisfaction with its appearance.

Several participants expressed frustration over having two sets of buttons that functioned as up-down buttons, not knowing which set to use. One of these participants opined, “I don't understand why there is an up/down button for warmer and cooler but also up arrows above the random buttons under the screen.” Another wrote, “I don't like that the menu button is between the warmer/cooler buttons and that it is the same shape & size.”

The videos, reviewed to shed light on the navigation problems, showed that some participants missed the "Accept" button when setting the target temperature, so their input was not saved. Others got lost in the different scheduling options offered because the names were so similar: quick schedule, program, or schedule.

5.4 Recommendations

Based on the results of this study, the research team makes the following recommendations for future programs that involve thermostats:

1. **Establish usability guidelines for thermostat purchasing decisions.** Such guidelines should indicate threshold scores for a subset of the metrics established in this study. At a minimum, the guidelines should take into account the thermostat features found to be strongly associated with task efficiency and thermostat preference scores. For example, these new purchasing guidelines might require all or some of the following:
    a. A high rating for feel and sound – e.g. at least a 5.0 on the 1-10 usability scale
    b. A high rating for ease of use – e.g. at least a 5.0 on the 1-10 usability scale
    c. A large screen – e.g. at least 4 square inches
    d. A color display – i.e. more than 2 colors

2. **Conduct a similar usability test for all thermostat models being considered for use in customer programs.** Thermostats that do not meet the established guidelines (see #1) should not be recommended, purchased by utilities, or installed in customer homes.

3. **Provide extra help for renters and the elderly.** If budget constraints disallow in-person, hands-on thermostat instructions for all customers, as much as possible, provide these services to renters and the elderly, who took significantly longer to complete the common tasks used in this study.

6 REFERENCES

California Energy Commission. *California Energy Demand 2010-2020 Adopted Forecast*. CEC-200-2009-012-CMF. December 2009.

Enerson, M. 2012. "Paired Comparison Usability Testing." Software Magazine. Available online at http://www.softwaremag.com/content/ContentCT.asp?P=3417 (accessed May 28, 2013)

Gunshinan, J. "Energy Star Changes Approach to Programmable Thermostats." *Home Energy*, March/April 2007.

Hair, J. F. Jr., R.E. Anderson, R.L. Tatham, and W.C. Black. 2009. *Multivariate Data Analysis*. NY: Macmillan.

Kennedy, P. 2008. *A Guide to Econometrics*. Oxford: Blackwell.

Meier, A., C. Aragon, T. Peffer, D. Perry, M. Pritoni. 2011. Usability of residential thermostats: Preliminary investigations. Building and Environment 46, pp. 1891-1898.

National Institute for Standards and Technology (NIST). 2001. Common industry format for usability test reports, version 2.0. Gaithersburg: NIST.

New York Times. *California Seeks Thermostat Control*. January 11, 2008. Available online at: http://www.nytimes.com/2008/01/11/us/11control.html.

Nielsen, J. 2005. "Time Budgets for Usability Sessions." Available online at http://www.nngroup.com/articles.

Norman, D.A. (1983). Design rules based on analyses of human error. Communications of the ACM, 26(4), 254-258.

Perry, D., C. Aragon, T. Peffer, A. Meier, M. Pritoni. 2011. Making Energy Savings Easier: Usability Metrics for Thermostats. *Journal of Usability Studies*, Vol. 6, Issue 4, pp. 226-244.

Rubin, J. and D. Chisnell. 2008. *Handbook of Usability Testing*. Wiley Publishing Inc., Indianapolis, Indiana.

U.S. Environmental Protection Agency. *ENERGY STAR® Program Requirements for Residential Climate Controls: Partner Commitments DRAFT 3 Version 1*. Available at https://www.energystar.gov/products/specs/node/161. April 9, 2012.