

How People Actually Use Thermostats

# Quantifying Usability in Programmable Thermostats

**Alan Meier** and Cecilia Aragon – LBNL

Therese Peffer and **Daniel Perry** – **UC Berkeley**

Marco Pritoni – UC Davis

\* [akmeier@lbl.gov](mailto:akmeier@lbl.gov)

# Thermostats

Premise: Improving the usability of thermostats  
will *facilitate* energy-saving behavior

# This Talk

1. Surveys of usability of thermostats in homes
2. Methodology for quantifying usability of thermostats and other controls

## Conclusions:

- Few homes exploit full potential of programmable thermostats
- It is possible to quantify usability in a way suitable

# Are Programmable Thermostats Used Correctly (or features fully exploited)?

Parallel investigations via:

1. Weatherization crews
2. Amazon Mechanical Turk
3. Interviews and other on-line surveys

# Survey of a Weatherized Home

- *“Do you use the programming features of the thermostat?”*

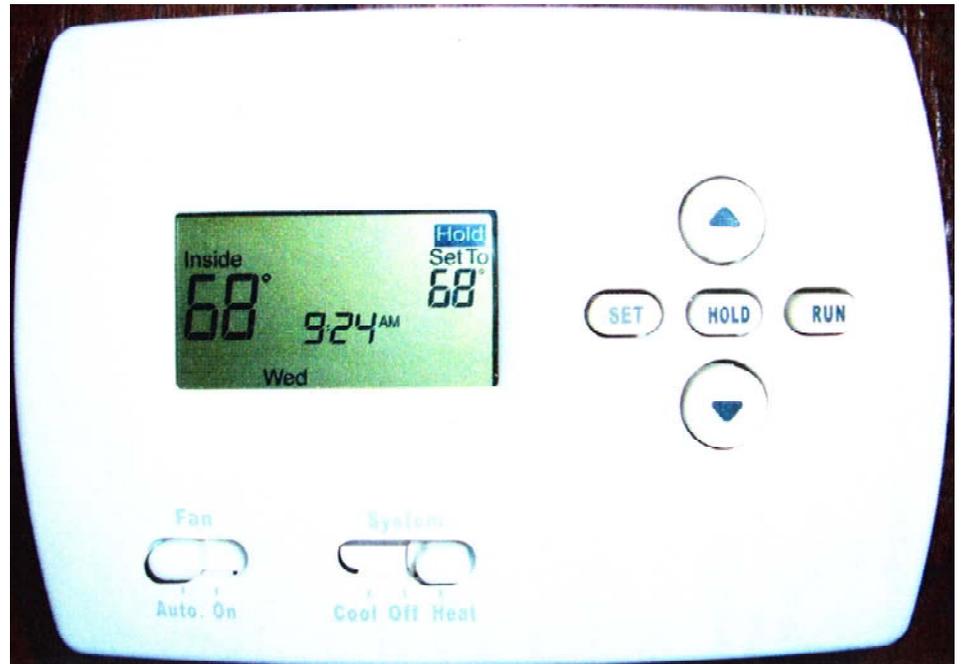
- *“Yes”*

- *“Are you satisfied with your thermostat?”*

- *“It’s OK”*

- *“If you could have a new thermostat, what would you like it to do differently from your current thermostat? (language, font size, button size, technical terms)?”*

- *“NO”*



***“The thermostat is on Hold”***

# Weatherization Survey Results

- 20 homes visited (in mid-west)
- 45% on “hold”
- 5% switched off (in winter)

# Amazon Mechanical Turk Survey

## Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

### As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

Fund your  
account



Load your  
tasks



Get  
results



## The Workers:

1. Fill out surveys on the web

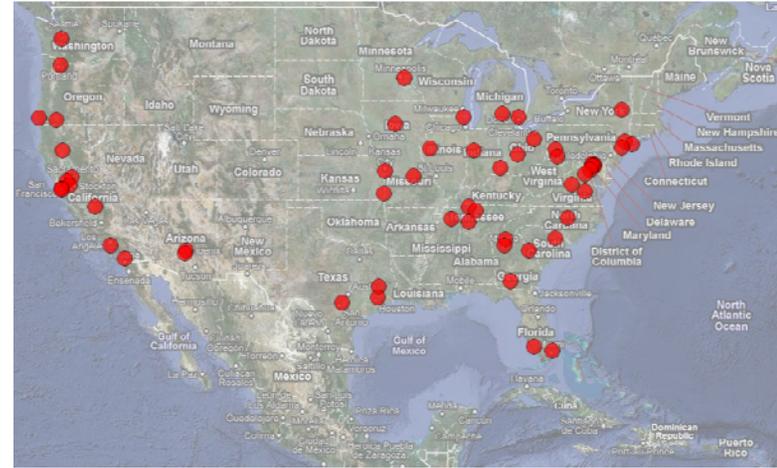
1. Photograph their thermostats

1. Get paid via Amazon

We get rapid,  
tabulated, results!

# Amazon Mechanical Turk Results

- 63 responses in 24 hr
- ~20% had errors in time setting
- ~50% on long-term hold
- Next survey will collect several hundred responses



Examples of uploaded photos

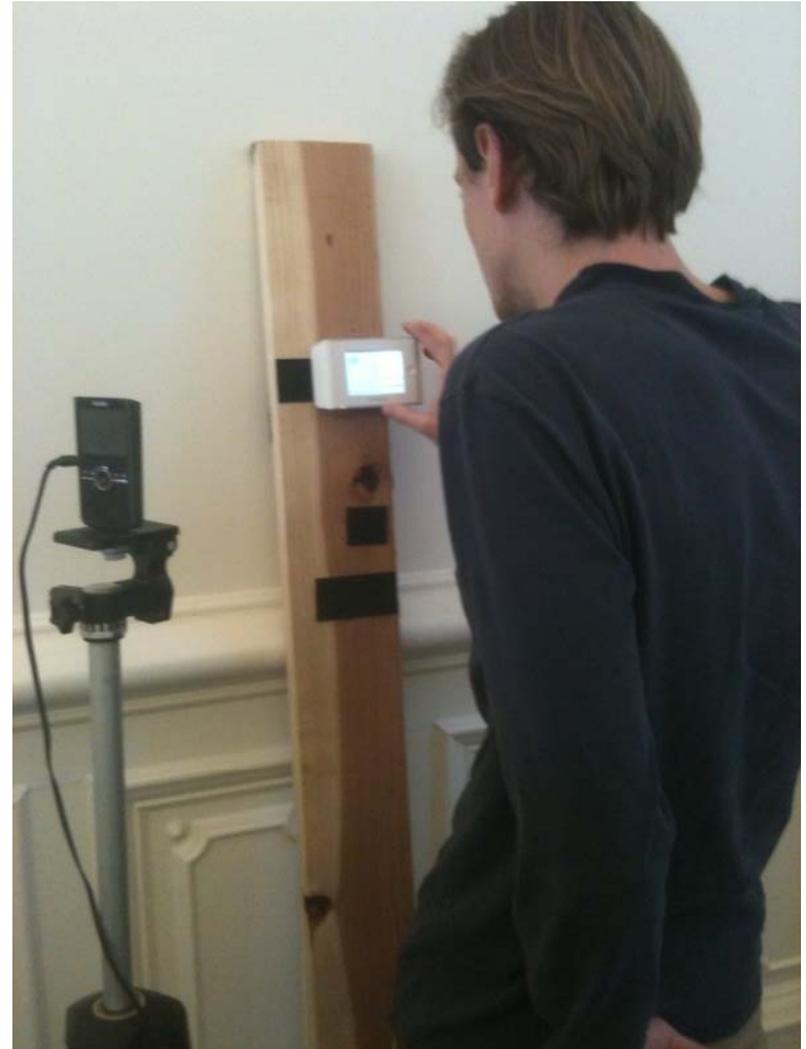
And Now the Video...

# How to Measure a Product's Usability?

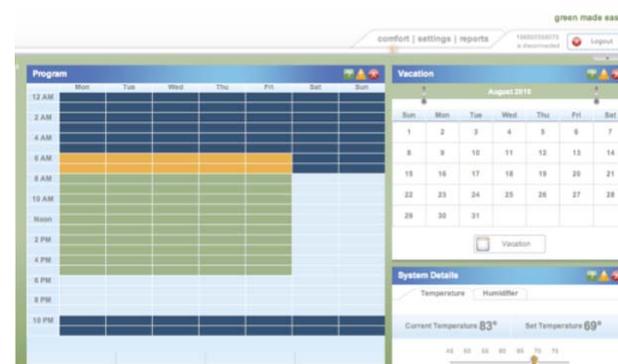
1. Define tasks
2. Quantify peoples' ability to accomplish tasks
3. Compute "score" based on metrics
4. Compare to reference model

# Details of Usability Tests

- 5 thermostat interfaces
- 31 participants
- 2 interfaces per person
- 6 tasks for each test
- 372 videos



# Thermostats Tested



# Define Tasks

**Task 1:** Turn the thermostat from “off” to “heat.”

**Task 2:** Set the correct time.

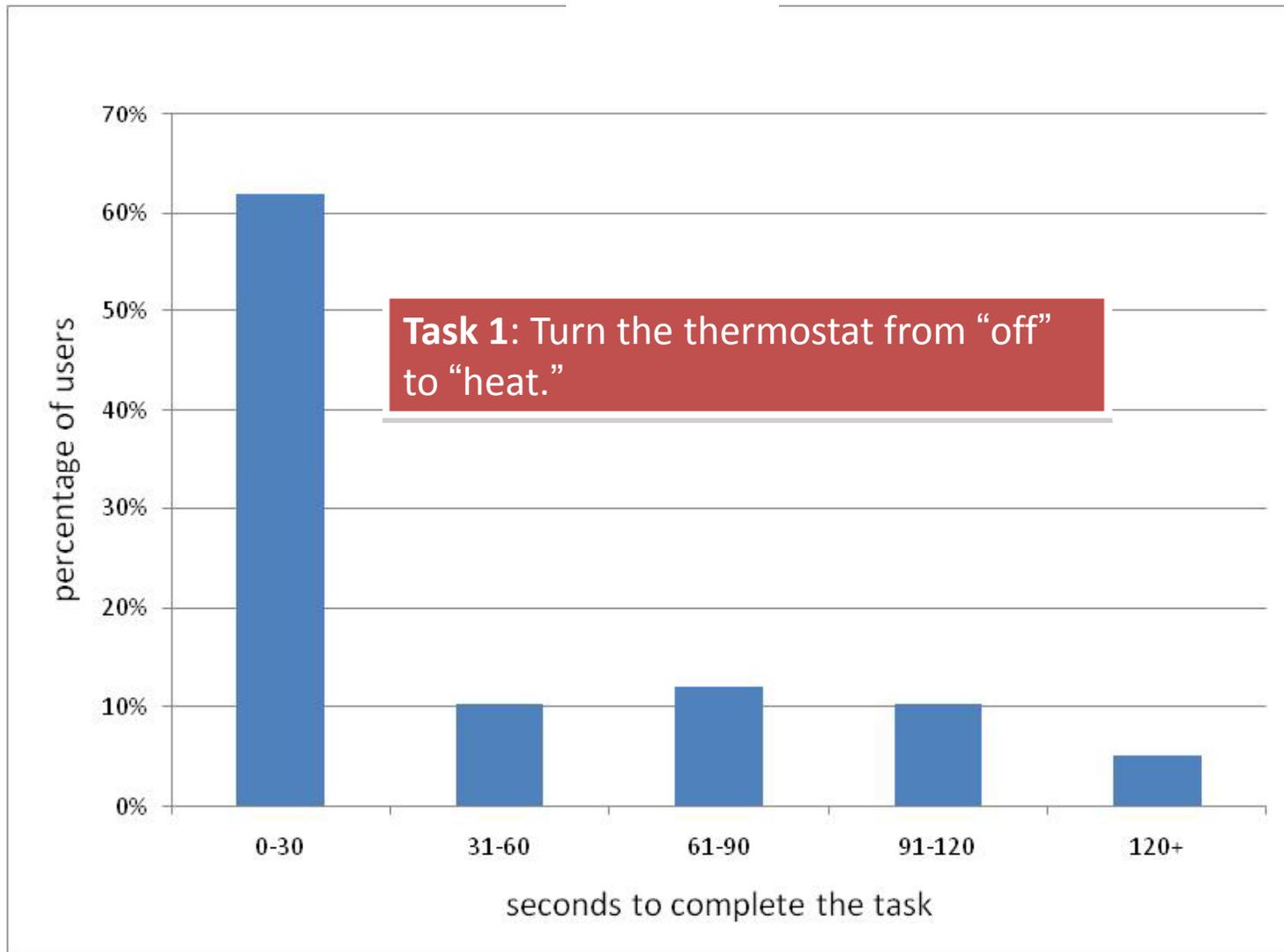
**Task 3:** Identify the temperature the device is set to reach.

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

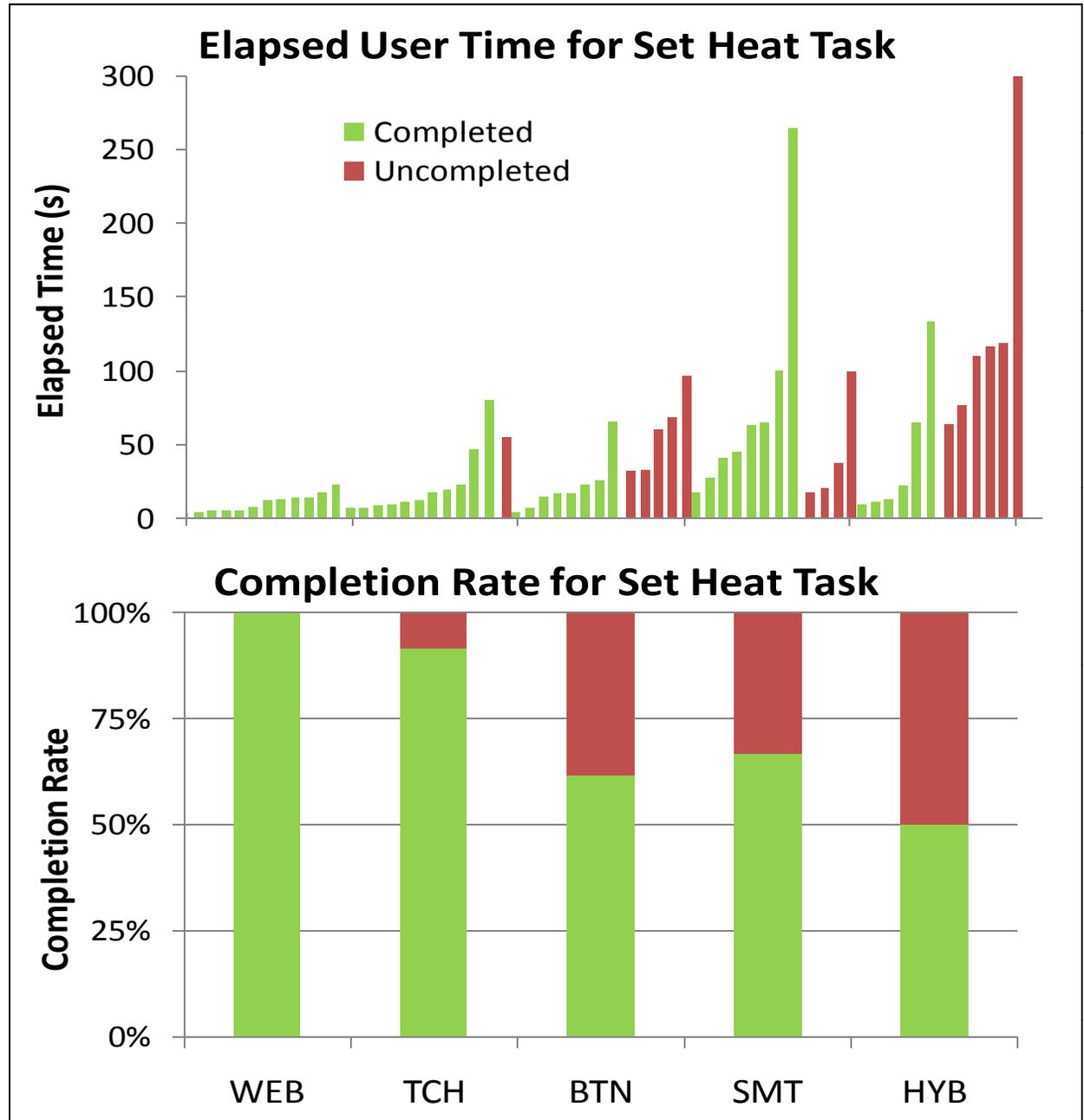
**Task 5:** Put the thermostat in “hold” or “vacation” to keep the same temperature while gone.

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

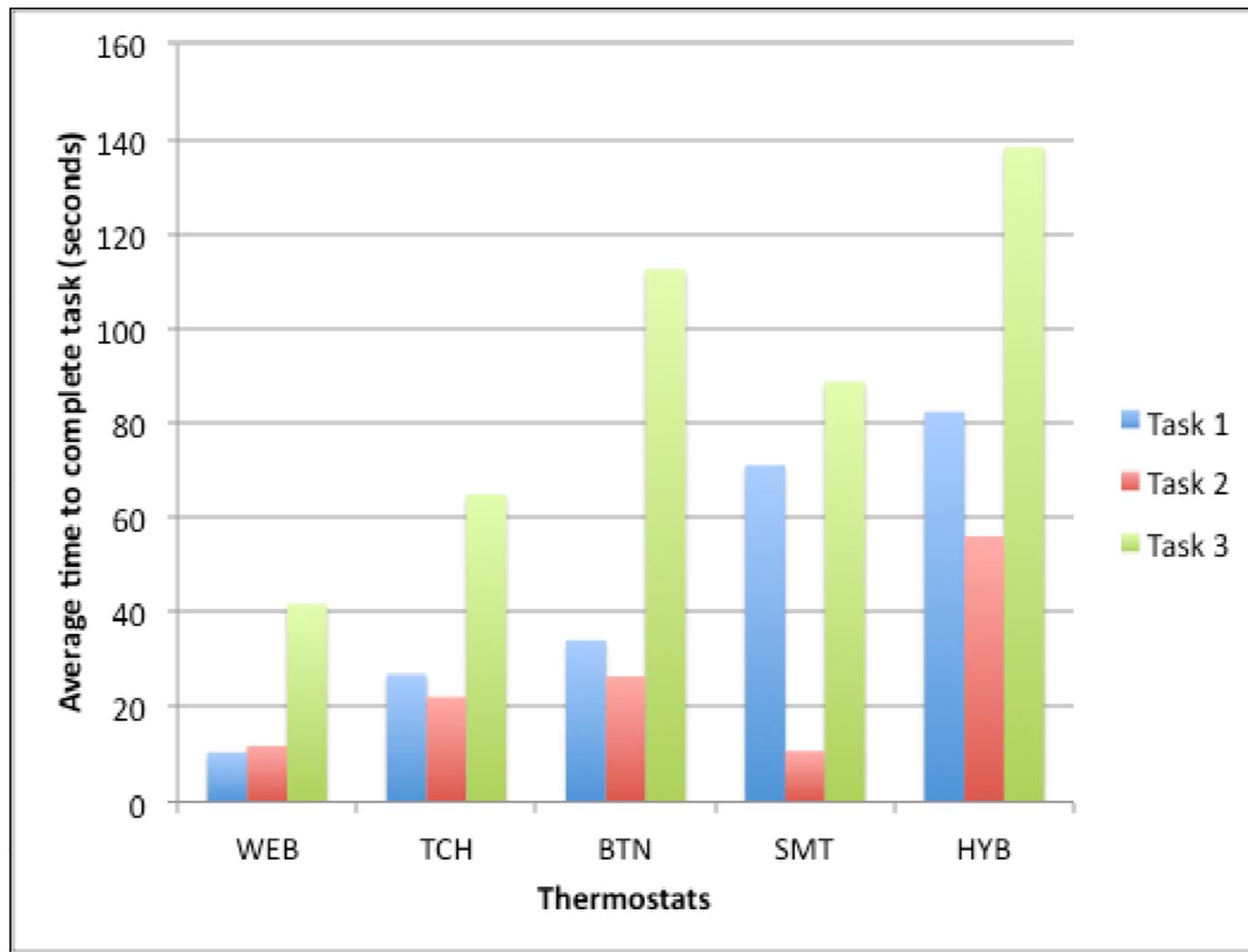
# Distribution of Times for Subjects to Complete Task 1



**Elapsed time and completion rate for subjects to accomplish the task of turning on the heat**



# Average elapsed time for subjects to accomplish Tasks 1, 2, and 3 with each thermostat



# Converting Videos of Tasks into a Usability Metric

# Quantifying Usability

- Many ways of translating observations in
- We created a procedure for normalizing data from different kinds of tasks
  - The procedure takes into account a subject failing to complete a task
- We examined four different metrics and compared results
  - How robust are results?

# Normalizing Data

We created a variant of the logistic function to normalize measurements so that all metrics would be between zero and 1:

$$P(x) = \frac{2}{1 + e^x}$$

# Taking Into Account Non-Completion

The formula is modified to account for non-completion of tasks.

The “ $M$ ” statistic is calculated for each metric  $i$  as follows on a per-trial basis:

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

where

$x_i$  = distinguishing variable for each metric

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

# The Efficiency Metric

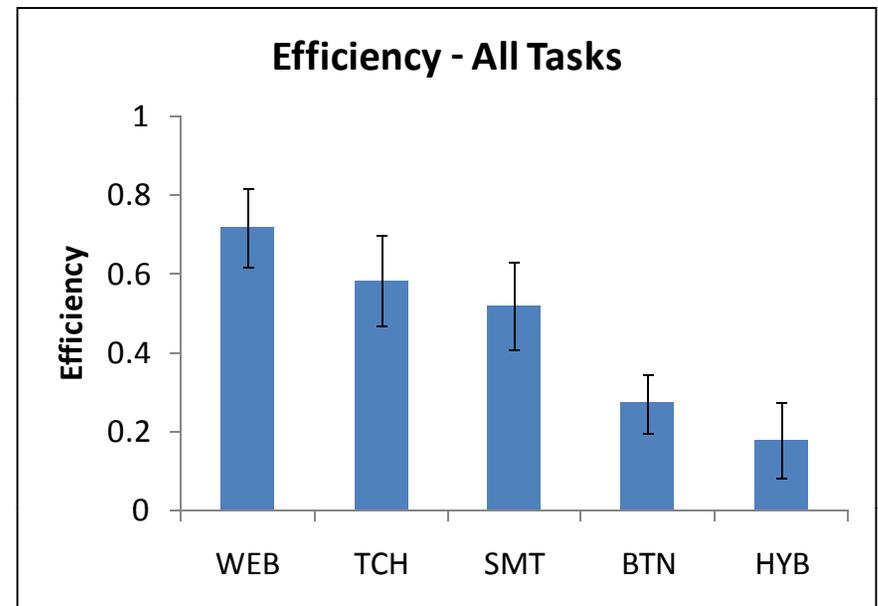
Time to complete task

$$x_1 = t/k_1$$

where

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

$k_1 = 50$  (empirically  
determined constant)



# Path Length Metric

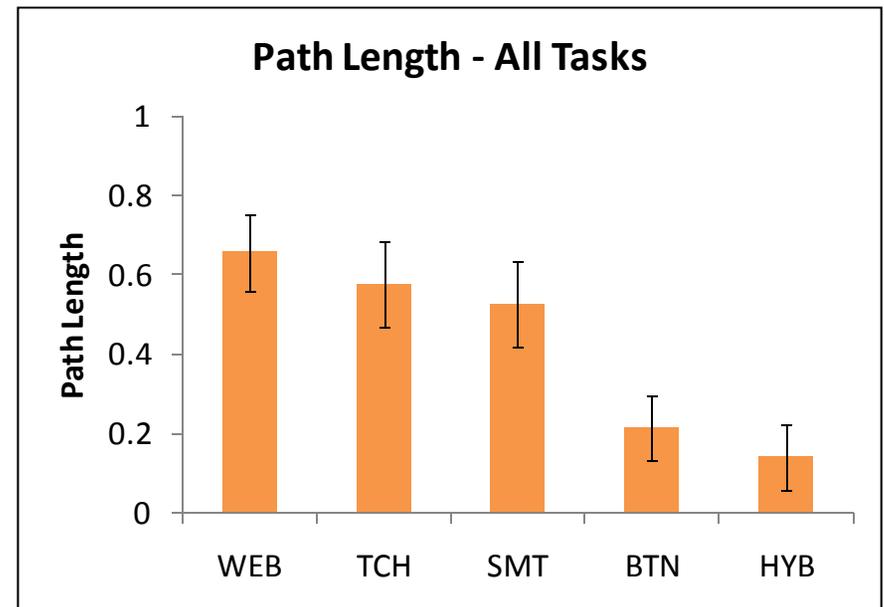
$$x_2 = \frac{f}{mk_2}$$

where

$f$  = number of buttons  
(functions) actually  
pressed

$m$  = minimum number of  
buttons (functions) required  
to complete the task

$k_2 = 5$  (empirically  
determined constant)



# Confusion Metric

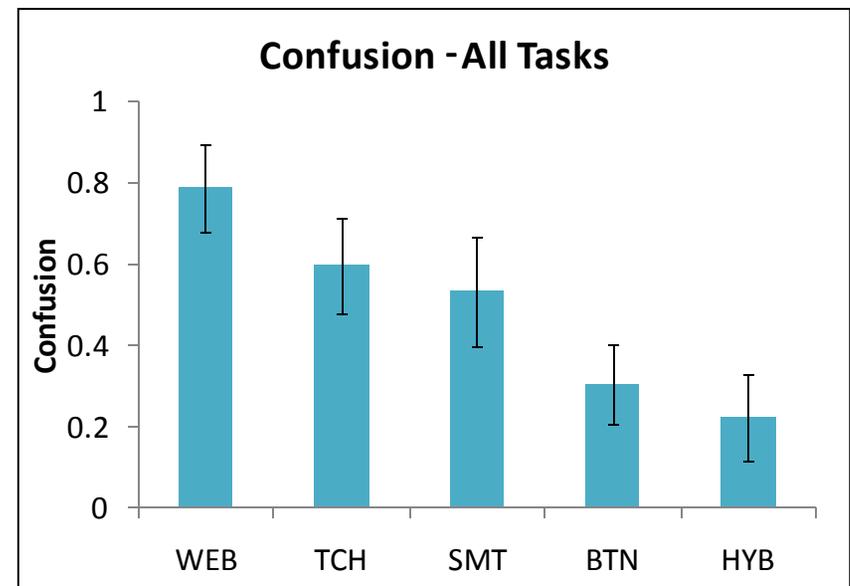
Sum of the time spent in hesitations,  $h$ , that users incurred over the course of a task. A hesitation was defined to consist of a pause or stop in user interaction for three seconds or longer.

$$x_4 = h/k_4$$

where

$h$  = sum total of durations of user hesitations

$k_4 = 2$  (empirically determined constant)



# Button Mash Metric

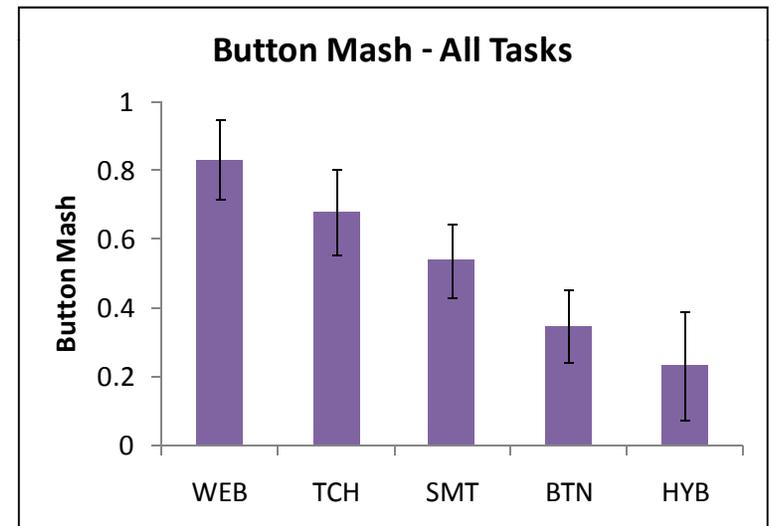
The sum of the number of times the user attempted to interact with the device but it had no effect.

$$x_3 = a_{ne}/k_3$$

where

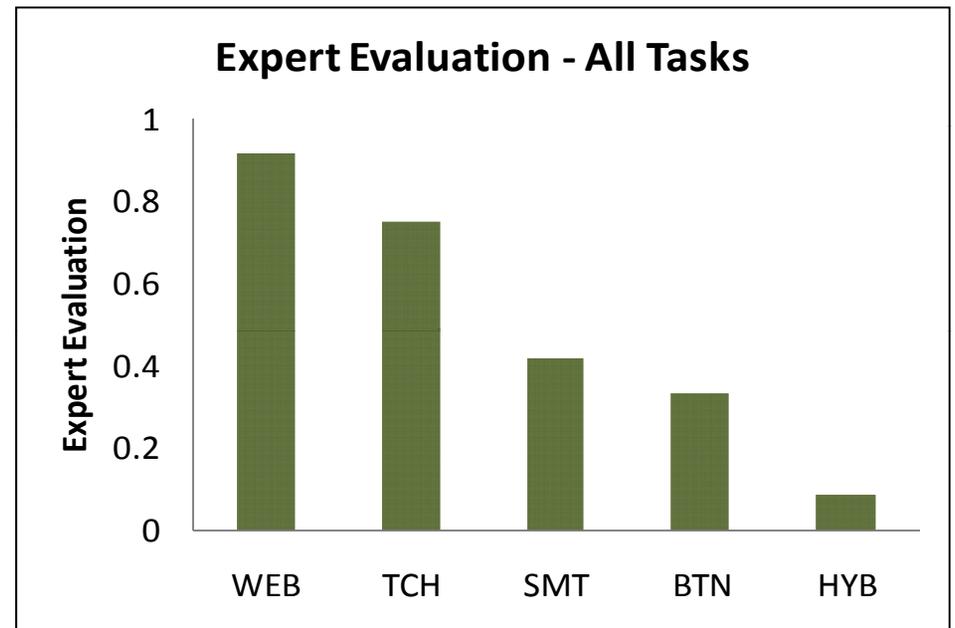
$a_{ne}$  = number of actions with no effect

$k_3$  = 5 (empirically determined constant)



# Expert Evaluation of Thermostats

Each thermostat underwent an expert evaluation to rate the usability of the device in performing the Set Heat, Current Settings and Future Settings tasks. These tasks were scored on a Likert scale of 1 - 5 where 1 was defined as "fairly easy" and 5 was "highly difficult" to use.



# Conclusions

- Field data suggest that energy-saving features captured
- It is possible to quantify usability of thermostat interfaces based on a series of representative tasks
- All metrics lead to almost identical rankings