Applied Energy 108 (2013) 448–458

Contents lists available at SciVerse ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Creating an in-home display: Experimental evidence and guidelines for design

Tamar Krishnamurti *, Alexander L. Davis, Gabrielle Wong-Parodi, Jack Wang, Casey Canfield

Department of Engineering and Public Policy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States

HIGHLIGHTS

- In-home display design approach incorporating customer preferences and experimental evidence.
- A computer-based simulated display to experimentally test feedback information.
- Contrast of customer feedback information preferences with experimental evidence.
- Appliance-specific/dollar feedback is not as effective as aggregated kW h feedback.
- Generalized information feedback may be more effective than a personalized display.

ARTICLE INFO

Article history:
Received 30 November 2012
Received in revised form 20 March 2013
Accepted 21 March 2013
Available online 18 April 2013

Keywords:
In-home displays
Electricity feedback
Consumer preferences
Learning

ABSTRACT

In-home electricity displays (IHDs) are digital devices that can give near-real-time information about electricity usage in the home. These devices have the potential to provide the kind of personalized feedback necessary to effect behavioral change among residential consumers. However, for consumers to be able to act on the information provided on IHDs, they must first be able to understand it. We present an approach to in-home display design that uses research on customer preferences to determine which features to experimentally examine for customer comprehension. Additionally, we compare these preferences against experimental data to determine whether people have insight into what information best works for an increased understanding of energy saving. Using a computer-based simulated IHD, we find that the types of feedback information that consumers prefer (appliance-specific and dollar-feedback) are not as effective for learning about appliance energy use as the less-preferred aggregated kW h feedback. Moreover, it appears that a simpler more generalized format of information provision has the potential to be more effective than a personalized IHD. We discuss how consumer preferences and experimental tests can jointly be used to inform the design of feedback technologies.

1. Introduction

Any information utilities give to residential electricity customers must be adapted to the customer's needs, especially for those who have limited knowledge of electricity-related concepts or a low literacy level. The monthly bill is typically the sole form of information provided to US households, and it is often too complex to be useful. Customers scan it to identify what they owe and then discard it without using the opportunity to learn—either because the information is not interesting or because it is not readily understood [1]. For example, electricity use information is typically presented only in kilowatt–hours (kW h), a unit that is opaque to many customers [2].

In the absence of usable information, customers will create 'folk theories' or mental models of how their appliances use energy [3–6]. If these theories are incorrect and people use them in their energy conservation strategies they may encourage waste, even with the best intentions [7,8]. Take, for example, the "valve" theory of thermostats, which holds that the quantity of cooling or heating in the home is directly proportional to the thermostat setting, rather than whether the setpoint is different from the current temperature. Those who believe this theory may set their thermostat very low (e.g., to 0 °C) hoping for faster cooling, only to waste energy when the air conditioner cools too much. Without information that corrects these folk theories, many customers would not understand how to adopt appropriate electricity-saving measures even if they wanted to.

Researchers and utilities have tried to solve this problem by providing customers with in-home electricity displays (IHDs) that can give near-real-time information about electricity usage. One of the earliest examples of a simple and particularly effective IHD was that used in the Twin Rivers study [9,10]. In this study, participants were given a simple light that flashed blue when one could cool the

* Corresponding author. Tel.: +1 412 445 2663; fax: +1 412 268 3757.
E-mail address: tamar@cmu.edu (T. Krishnamurti).

0306-2619/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apenergy.2013.03.048

home by opening the windows rather than using the air conditioner. This yielded an almost 20% reduction in monthly electricity use over the short duration of the study. Paired with the digital meters ('smart-meters') of the smart grid, more sophisticated IHDs can provide customers much higher resolution feedback about their electricity consumption. If this feedback is presented in the right way, customers should be able to correct their mental models of how appliances use energy in much greater detail, allowing them to more easily engage in energy efficient behavior.

Since the Twin Rivers study, mounting evidence has shown that IHDs can help customers curtail their electricity use. A variety of displays have been used in these field studies, including retail (e.g., the PowerCost Monitor) and custom devices (e.g., The Residential Energy Cost Speedometer; [11]). Each display provides different types of feedback information (e.g., kW h use, cost of electricity, monthly spending), in different formats (e.g., graphs, tables, numbers, visual-analogs). In a recent review of these field trials, Davis et al. [12] found that four custom displays (Bluelight, [9]; RECS [11], Fitch [13] and Electricity Consumption Display [14]) were the most effective for reducing overall consumption (~20%, ~13%, ~12%, and ~12%, respectively). It appears that custom designed IHDs can provide the right information in an easily understood manner, leading to effective reductions in electricity use.

While these findings are encouraging, the small sample sizes of these studies (N = 20, 99, 101, and 8, respectively) alone should raise doubts about their real-world effectiveness. Casting further doubt, field studies of IHDs report methodologies and measurements that vary so much it is difficult to quantitatively aggregate them, or even compare them to one another [15–19]. While studies that demonstrated large versus small effects differed in many ways, an important difference was the type of IHD they used, suggesting that specific features of the displays may play a unique role in spurring energy reducing behavior.

1.1. Beyond preferences

To determine whether features of the display matter, one can just ask customers what they want, as they have strong preferences about the kinds of information they want to see. Relying on preferences alone to infer how customers will behave, however, is an incomplete approach, as "what people think they want and what they actually want are not always the same" [20]. Basic psychological research has shown that people are not always good at predicting what they will like, concentrating too much on changes [21] or showing bias toward their present feelings [22]. People also have been known to reject policies in prospect, but like them once implemented [23]. Thus, examining preferences alone may give a certain, but potentially incorrect, perspective of how an in-home display should be designed to be most effective.

Beyond their preferences, a variety of social factors will certainly affect consumers' ability to translate the information they view on the display into actual behavioral change. The novelty of the experience [24], willingness to conserve [25], household disposable income [26], cultural norms of energy savings [27] and physical limitations, such as lacking the ability to repair or replace inefficient appliances (e.g. low income public housing residents, see [28]) are just some of the factors that can affect whether feedback information is effective.

In this paper, we take a step back from real-world use of IHDs to examine the more basic question of whether consumers can actually understand and learn from the individual types of feedback information they might see on an IHD,[1] using a simple computer-based in-home display simulation. By identifying those features that best allow for learning, we can begin to make recommendations about which features to include on an IHD to potentially prompt behavior change.

Our approach complements this research by using consumer preferences to determine which features to experimentally test. We then compare preferences against experimental data to determine whether people can use the kind of feedback information that they believe would allow them to change their behavior. To date, little experimental work (with the exception of enhanced bills, [29,30]) has been conducted. The various field studies, interviews, and surveys have neither separated specific elements of IHDs according to their effectiveness, nor measured important intermediaries of effectiveness, such as learning and motivation [31–35].

1.2. Existing research on consumer preferences

Past research on customer preferences for IHD features has used interviews, surveys, and other similar approaches (e.g., focus groups). The options participants generated or could choose from have generally fallen into five categories outlined below [32,36].

1.2.1. Units

Information about electricity can be displayed on an IHD in different units, such as current cost ($), cost/day, power (W or kW), energy (kW h), or carbon dioxide emissions (CO_2 tons). In general, people prefer the cost of electricity above all other possible ways to display electricity use [36,20]. This is consistent with customers wanting simple information in units that they already understand. A number of studies have found that people prefer seeing their costs either as current rate of expenditures (in $/day) or cumulative cost in $ per billing period [36].

1.2.2. Time aggregation

Information can be displayed in increments ranging from years to real-time updates. Unlike preferences for units, there appears to be no consensus regarding preferences for time aggregation. Some, for example, prefer to see their electricity consumption on an hourly basis [20], while others prefer to see it on a quarterly basis, compared to some reference point like the previous quarter [31]. Still others prefer to see their electricity use displayed as daily load curves [37] rather than 10-day curves [38]. However, while there is no unanimous preference for time-period, people generally want to be able to switch time periods with the press of a single button [2,20]. Although monthly billing information is common, more frequent information may be helpful [24].

1.2.3. Physical aggregation

While we know of no research on whether people prefer electricity use information by room, by specific household member, or for the whole house, two recent studies found that people strongly prefer appliance-specific information [36] in monetary units [20].

1.2.4. Comparators

Comparisons typically examined have been to oneself (historic), to other customers (social), or to targets (goal). The most frequent finding is that people want to compare their current use to their own use at some point in the past [31,36]. Moreover, people want to compare their personal electricity use to a self-set goal or target [2,20]. In contrast, nearly all people express a strong rejection of social comparisons [31,20,39], wherein they see their electricity use compared to some other group of customers, such as their neighbors. Indeed, there is little evidence suggesting social comparisons motivate people to reduce their household electricity use [33].

[1] See Wilhite and Ling's 'information-deficit' model for a lengthier discussion of why knowledge is a crucial precursor to behavior change [24].

1.2.5. Format

Displayed information can be formatted as a chart, picture, table, numerically, as text, or as a combination of audio and visual feedback. Like time aggregation, there is a lot of variability in preferences for format, with some preferring visual-analog 'speedometers' [20], and others preferring bar graphs or bell curves [34]. Karjalainen [36] developed eight paper-based IHD prototypes and evaluated them using preference assessments and think-aloud protocols. People understood bar charts, pie charts, and numerical tables easily. However, the researchers found that a tabular display was not only, "instantly understood by everyone," but also preferred the most, as $\frac{7}{15}$ participants ranked it the best of the eight prototypes.

Overall, the research shows some agreement on what customers want in terms of units and comparators, while there are fewer findings on time aggregation, physical aggregation, and format. There is also evidence on customer preferences for esthetics of information (see, for example, Paetz [39] for a discussion of participant preferences on color or Wood and Newborough [7], for a discussion of size and legibility of text).

In the following section we present new data on customer preferences for the attributes found on the most common commercially available IHDs. We then present findings of an experimental test of two units of information ($ versus kW h information) and two forms of aggregation (total versus appliance-specific), using an IHD simulation. We conclude with a discussion of the implications of our findings, as well as how our approach can inform the development and testing of new feedback technologies in the field.

2. Survey

2.1. Materials

To choose which IHD features to test we developed a list of 19 displays from a larger set of those that were commercially available to residential customers. We then developed a list of the most common types of electricity feedback information either provided by these 19 displays or that were preferred in previous literature on IHDs.

2.2. Procedure

Participants were asked to rate this set of feedback information with the following instructions: "Here is a list of information that might appear on an in-home display. Please rate each type of information in terms of how much you would like to have it on the display." Each feedback information type was rated from 1 (not at all) to 5 (extremely), with the additional option of responding "I don't know".

Table 1
Feedback information preferences.

Feedback information type	Mean	SD
Bill-to-date	4.11	0.94
Appliance-specific	4.00	0.96
Daily projections	3.83	0.99
Monthly projections	3.80	0.96
kW h -to-date	3.80	0.96
Daily price	3.70	1.16
Daily peak use	3.69	1.04
Monthly peak use	3.68	1.08
Goal tracking	3.61	0.96
'Greenness' of use	3.40	1.10
Usage comparison to a similar household	2.87	1.20

Participants then created their own display by selecting which features they wanted from the same list of feedback information types. They also responded to the following questions about the display they created, "How much would you like to have the in-home display you created on the previous page? (1 = not at all; 5 = extremely)," "How effective do you think that in-home display would be in helping you to reduce your electricity use? (1 = very ineffective, 7 = very effective)," "How often would you look at the in-home display you created? (1 = never, 7 = daily)," and "How much do you think you would save, in dollars, on your monthly electricity bill if you had the display you created on the previous page?"

2.3. Participants

Participants were bill-paying electricity customers ($N = 151$) in the United States recruited using the Amazon MTurk system.[2] Their average age was 32 years old ($SD = 11$ years), with 42% being male, and most having an income between \$51 K and \$75 K per year. The average electricity bill among these customers was \$106/month.

2.4. Results

2.4.1. Feedback information preferences

Table 1 presents the feedback information types in order of preference. A Wilcoxon signed-rank test was used to assess whether participants ranked one attribute higher than another. Information presented as "bill-to-date" and "appliance-specific" feedback were considered the most desirable IHD features, although neither was preferred over the other according to the Wilcoxon signed ranks test ($Z = 1.6$, $p = 0.11$). The least preferred way of presenting information was a comparison to a "similar household," which was rated as much less desirable than any of the other attributes ($p < 0.001$ for all comparisons).

2.4.2. Create-your own

Participants constructed IHDs that were almost identical to their ratings, so we omit the details of their preferences here. On average participants strongly expected to like the display that they created, as seen in the mean ratings of expected liking being significantly above the scale midpoint of 3.0 in a one-sample t-test, $t(137) = 13$, $p < .001$. They also thought it would be an effective way for them to reduce their electricity use, again with mean ratings of expected effectiveness being significantly above the scale midpoint of 4.0, $t(138) = 21$, $p < .001$, and expected to look at it an average of 2–3 times a week ($M = 6.11$, $SD = 1.1$, $N = 138$). They anticipated an average savings on their monthly bill of \$25 ($SD = 29$, $N = 131$).

2.5. Discussion

Participants reported the strongest preference for bill-to-date and appliance-specific feedback. This finding is consistent with previous research on preferences for IHD content [2,36,20]. For the most part, they did not differentiate between the other types of information, such as projections or goals, but they did report a strong dislike of the more 'gimmicky' features, such as greenness and social comparisons. Expectations for monthly savings from

[2] Amazon MTurk participants have been shown to provide reliable data, replicating results across a variety of psychological and economic studies, including a number of classic judgment studies [40,41]. Additionally, MTurk participants will work at a high level of accuracy for payment levels well under those that we provided [42]. MTurk samples also provide a greater degree of demographic diversity than is available from other convenience samples (i.e. internet panels or college undergraduates). Furthermore, when responding to measures, they give answers that have as good psychometric properties (e.g. internal consistency) as other internet samples [43].

Fig. 1. Screenshot of the simulated in-home display.

an IHD were unrealistically high (about 25% of their monthly bill), given that the average monthly savings from IHDs have peaked at 15–20% in previous research [12,18,19,26].

Our survey confirms findings from previous research on customer preferences for feedback information showing that consumers want to be able to look at an IHD and quickly determine how to manage their monthly budget (bill-to-date) and control their current consumption (appliance-specific feedback). Our next study draws on these preferences to test how well consumers learn from what they say they want. Specifically, we test whether they can learn from simple bill-to-date and appliance-specific feedback, and if it is possible to get the same benefit using only kW h. The experiment described in the next section addresses this question using a computer-based simulated in-home display.

3. Simulated in-home display experiment

3.1. Methods

Participants interacted with a computer-based in-home display simulation that allowed them to turn eleven appliances on and off, change the settings on thermostats for various appliances, and alter how long to run each appliance (in 30 min increments), imagining they were in the house of a fictional family, "The Smiths." Feedback information on electricity use was updated according to these manipulations and presented in a tabular format (see [36] for a justification of this choice), as seen in Fig. 1.

We selected the eleven most commonly owned electric home appliances in the US: (1) air conditioner, (2) water heater, (3) indoor lights, (4) outdoor lights, (5) refrigerator, (6) freezer, (7) oven, (8) microwave, (9) television, (10) washing machine, and (11) dryer. We used standard estimates for the real-time electricity consumption values[3] and the standard estimate for the average cost of electricity in the US of \$0.13 per kW h.[4] When cost and usage feedback was provided to the participants on the simulated IHD, these standard values were used to calculate that feedback.

3.2. Experimental conditions

Participants were randomly assigned to one of six in-home display simulations that provided them with different feedback. The specific conditions were designed to test whether participants could learn from the information that people preferred (appliance-specific feedback in \$ units) versus the information that would be easiest for a utility to provide (aggregate kW h usage). Details of each condition can be seen in Table 2. Learning was assessed on ten out of eleven appliances.[5]

We also included a control condition simulating more generic educational materials that could be provided on a bill insert. In this *passive learning* condition, participants were simply provided with information on how much energy each appliance uses and how much each contributes to the monthly bill [7]. This type of passive, as opposed to discovery, learning can come from a variety of sources such as "educational campaigns, advertisements, advisory services and news media" [2]. Both theory and practice suggest that passive learning is not sufficient for behavior change, although it may be more effective in cases when behavior change is convenient and cheap (low barriers) [23]. Here passive learning serves as a control by establishing whether participants can process the information that they were given, knowing that aspects of this topic may be something that participants have had little or no prior exposure to.

3.3. Participants

Participants were bill-paying electricity customers ($N = 191$) recruited using the Amazon MTurk system. Forty-two percent were

[3] http://visualization.geblogs. com/visualization/appliances.
[4] http://www.eia.gov/electricity/monthly.

[5] Although participants were able to manipulate and received feedback on the electricity consumption of 11 appliances, pre-testing showed that participants felt much more comfortable with 10-item rankings. The difference in usage between indoor and outdoor lights were marginal compared to the other appliances, therefore outdoor lights were removed from the ranking tasks.

Table 2
Description of feedback information for the six conditions and one control group.

Group condition	Feedback information
kW h only	Participants were shown kW h feedback information used for running the appliances at the settings they made for the duration they specified
$ only	Participants were shown cost ($) feedback information used for running the appliances at the settings they made for the duration they specified
kW h and $	Participants were shown both kW h and cost ($) feedback information used for running the appliances at the settings they made for the duration they specified
$ by appliance	Participants were shown cost ($) feedback information in total, as well as for each appliance, used for running the appliances at the settings they made for the duration they specified
kW h by appliance	Participants were shown kW h feedback information in total, as well as for each appliance, used for running the appliances at the settings they made for the duration they specified
kW h and $ by appliance	Participants were shown kW h and cost ($) feedback information in total, as well as for each appliance, used for running the appliances at the settings they made for the duration they specified
Passive learning	Participants were given the answers to the electricity knowledge pre-test and then asked to provide those answers in the post-test

male. Participants ranged in age from 18 to over 65 years, with most participants between 22 and 34 years old. Most participants had an income of $51–$75 K per year. The average electricity bill among these customers was $138/month.

3.4. Measures

Participants were provided with a basic electricity knowledge test (some questions adapted from [44]) before and after interacting with the simulated IHD. This test assesses understanding of the concepts that could be learned from the simulation. It also measures the basic knowledge needed to be able to interact with the display (e.g. what a kW h is), as well as an understanding of which appliances may contribute most to problematic energy use.

Specifically, the test assessed participants' rankings of ten appliances in terms of kW h consumption in a 10-min period (i.e., power consumption) and contribution to the monthly bill (i.e., cost of energy consumption). For example, participants were asked to "Imagine each appliance listed below is used for exactly the same amount of time (10 min). Rank the ten appliances below by how much electricity they use from 1 (the most) to 10 (the least)." Participants were also asked to answer three basic conceptual questions: (1) the units of electrical energy, (2) the correct method of calculating energy, and (3) the estimated the cost of a kW h (we used $0.13 as an approximation). All measures and the simulation were pre-tested with nine in-depth cognitive interviews.[6] The interviews were used to ensure that the simulation was comprehensible, usable, and that the questions were interpreted as intended.

3.5. Procedure

Participants completed the ranking questions (pre-test) prior to viewing the IHD simulation. They were then randomized to one of the seven conditions previously mentioned. Interaction with the simulation (and viewing the passive learning information) lasted for as long as they wanted. After interacting with the simulation, they completed the same ranking questions (post-test). Lastly, they completed demographic questions.

3.6. Results

3.6.1. Level of Interaction

Participants spent a median of 34 min (SE = 3.1 min) interacting with the IHD simulation. This interaction time did not differ by conditions (all t-tests were less than 1.4, not reaching statistical significance).

[6] Lab notebooks for these interviews are available at: http://openwetware.org/wiki/IHD_Simulation.

3.6.2. Appliance rank deviations: power consumption

Figs. 2 and 3 show pre-post differences in rankings of appliances by hourly power consumption, aggregating across conditions. As can be seen from the histograms, with bolded lines indicating the true ranking of the appliance and dashed lines indicating statistically significant shifts at post-test, deviations from the true rank improved for AC, Microwave, and Fridge ($Z = 2.6$, $p = 0.01$; $Z = 4.8$, $p < 0.001$, and $Z = 3.1$, $p = 0.0021$, respectively). There were no overall differences for the dryer, oven, water heater, washer, freezer, TV, and indoor lights, ($Z = 1.6$, $p = 0.11$; $Z = 0.043$, $p = 0.97$; $Z = 1.1$, $p = 0.27$; $Z = 1.4$, $p = 0.16$; $Z = 0.7$, $p = 0.48$; $Z = 1.9$, $p = 0.054$; and $Z = 0.63$, $p = 0.53$, respectively).

Participants were very accurate in ranking the highest and lowest use appliances, the AC and indoor lights, both before and after interacting with the simulated IHD. Those appliances that participants were initially least sure about (as illustrated by the scattered nature of their initial rankings), the water heater and the oven, did not improve after interacting with the simulated IHD. The most striking improvement was for the microwave, an appliance that participants had certain but incorrect initial beliefs about.

3.6.3. Appliance rank deviations: cost

Figs. 4 and 5 show pre-post differences in rankings of appliances by monthly contribution to electricity bill, aggregating across conditions. Results for cost rankings are similar to those for power consumption. Participants improved from their initial rankings across conditions for the washer and freezer ($Z = 3.4$, $p < 0.001$ and $Z = 3.5$, $p < 0.001$, respectively). Their rankings at post-test were worse than their pre-test rankings for the oven, water heater, and microwave ($Z = 0.28$, $p = 0.78$; $Z = 1.6$, $p = 0.12$; $Z = 1.1$, $p = 0.29$, respectively). For the AC, lights, TV, Dryer, and Fridge, participants were initially quite accurate in their rankings and either remained accurate or slightly improved ($Z = 0.029$, $p = 0.98$; $Z = 0.59$, $p = 0.56$; $Z = 1.6$, $p = 0.12$; $Z = 1.1$, $p = 0.29$; $Z = 0.62$, $p = 0.54$, respectively).

Contrary to the findings for power consumption, the post-simulation rankings for the microwave were more inaccurate than the pre-simulation rankings. However, unlike their precise but incorrect initial beliefs about how many kW hs it used in 10 min, participants had imprecise initial beliefs about how much the microwave would cost them in a month. Participants learned that the microwave used much more power (kW h) than they expected, but they seemed to incorrectly extrapolate this greater power use to monthly energy cost (i.e., kW h cost). This confusion stemmed from a failure to take into account how infrequently the microwave is used compared to other appliances.

3.6.4. Appliance rank deviations by specific condition

Following the aggregate analysis, we looked at how the rankings of each appliance changed for each treatment condition individually. Hierarchical linear models were used to examine the

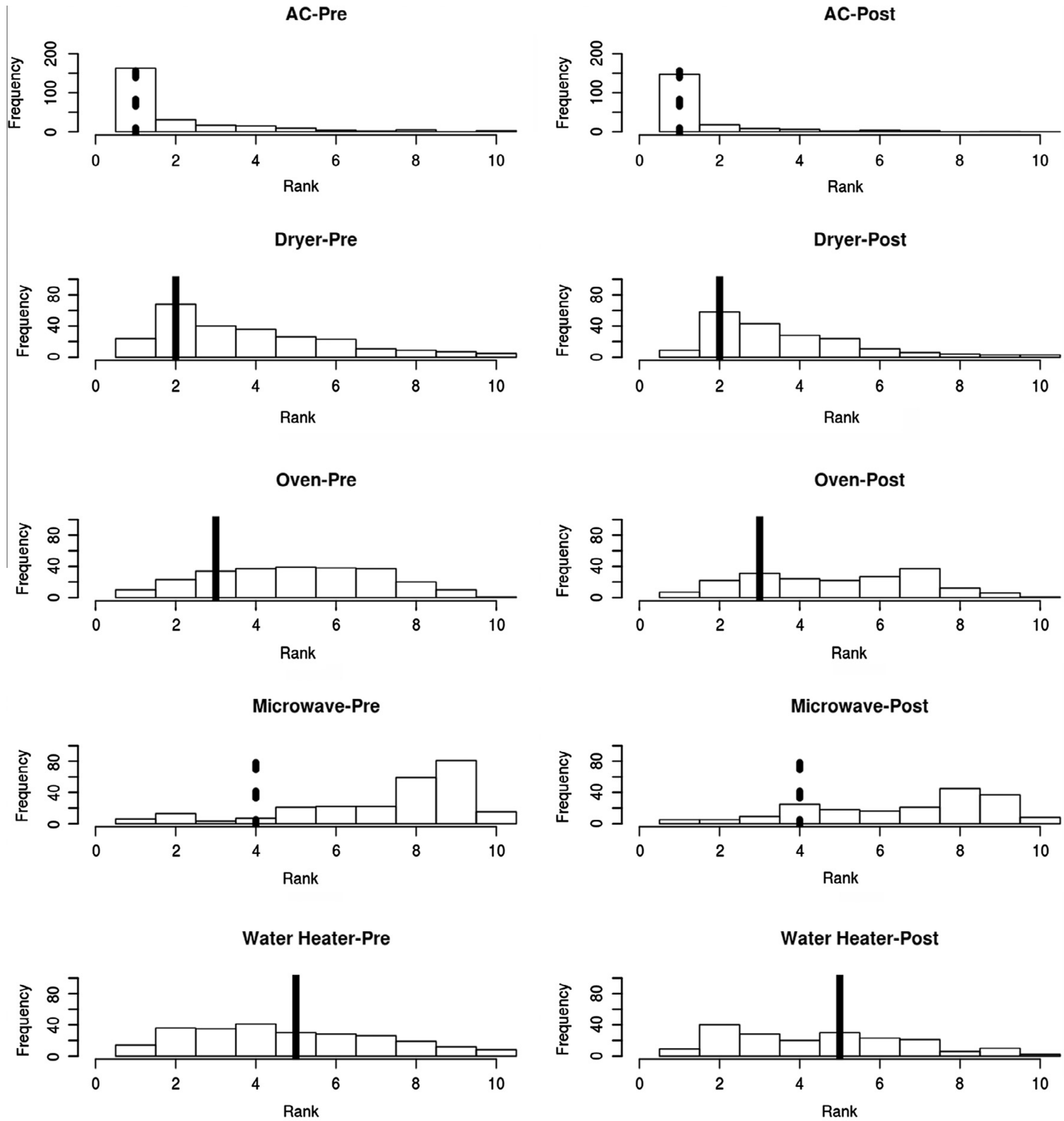

Fig. 2. Histogram of kW h use rankings for each appliance in order of true ranking. Lines indicate true rank, with dashed lines indicating statistically significant pre-post differences.

effect of condition on learning how much power each appliance uses and how much each appliance contributes to the monthly bill [45,46]. The results of these models are shown in Tables 3 and 4, respectively. Each model uses different intercepts for each appliance and examines the effect of each treatment condition aggregating across appliances, weighting them appropriately.

Although not statistically significant, those given kW h information were more accurate in their rankings of how much power appliances use, but not how much each appliance contributes to the monthly bill. Conversely, and again not statistically significant, those given $ information were more accurate in their rankings of how much each appliance contributes to the monthly bill, but not how much power each appliance uses. There was no benefit of having appliance-specific information either for learning how much power each appliance uses or how much they contribute to the monthly bill.

3.7. Basic conceptual knowledge

As seen in Table 5, in all conditions participants were more likely to change their answer from an incorrect response at pretest regarding the definition of a unit of energy to the correct one (kW h) at post-test than from correct to incorrect. However, none of the differences were statistically significant. Aggregating across all conditions, participants were more likely to change their answer in the correct direction. As there is no control group of participants who did not receive any feedback (i.e., merely completed the pre and post questionnaires) to compare them to, this positive change could be an effect of learning from those questions that frequently asked about kW h, rather than from the feedback alone.

As seen in Table 6, participants were not more able to infer how energy is calculated at post-test in any of the treatments. Most

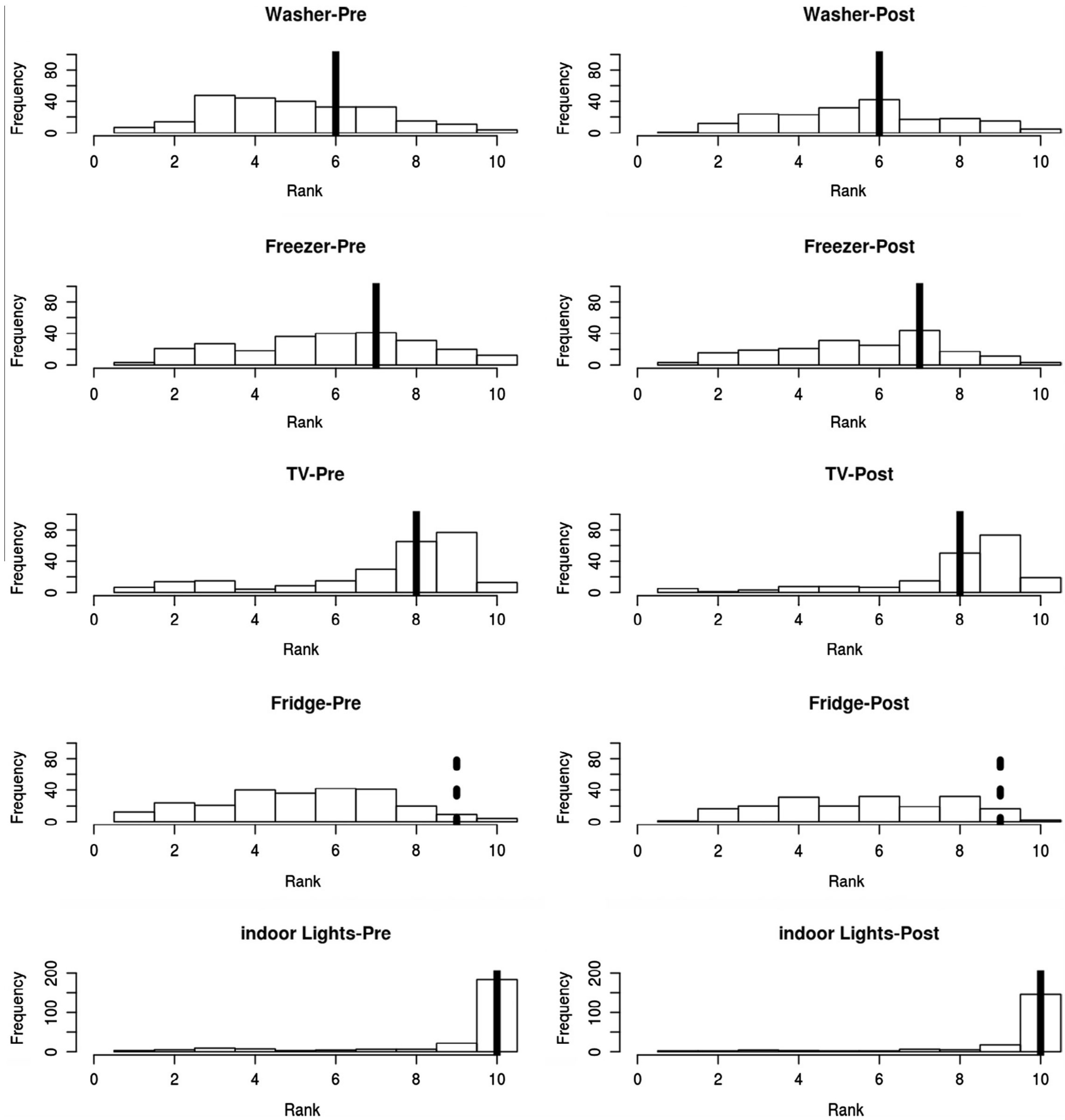

Fig. 3. Histogram of kW h use rankings for each appliance in order of true ranking. Lines indicate true rank, with dashed lines indicating statistically significant pre-post differences.

(136) participants did not change their answer from pre-test. Some (34) changed from the wrong answer to the right answer, whereas others changed from the right answer to the wrong answer (22). There was no detectable variation by condition, but aggregating across all conditions there was improvement.

As seen in Table 7, all participants who received $ information provided responses closer to the true kW h cost of $0.13 in the post-treatment period. On the other hand, participants who were provided only kW h information (kW h by appliance and aggregate kW h only) were unable to learn the true cost of a kW h, showing no improvement. The only statistically significant improvement from pre to post-test was for participants in the passive learning condition, the median of which got the true kW h cost exactly correct. Additionally, those given some form of $ information also showed they could learn even though the *p*-values do not reach significance, likely due to the small number of non-tied (N_t) deviation scores in these conditions.

3.8. Discussion

When participants had preconceived (but incorrect) notions of how much power an appliance used, as in the case of the microwave, they were able to correct their beliefs by learning from the simulation. Learning was more modest or non-existent for appliances that they knew little about beforehand (the water heater and the oven), as evidenced by varying opinions between participants about the rankings of these appliances both before and after interacting with the simulation. Even without interacting with the simulation, participants demonstrated high accuracy in identifying which appliances used the most (air conditioner) and least (indoor lights) out of the ten.

Learning did not depend on whether participants were provided with appliance-specific feedback, even though they often referenced specific appliances when providing open-ended explanations of their performance at the end of the task. They may have

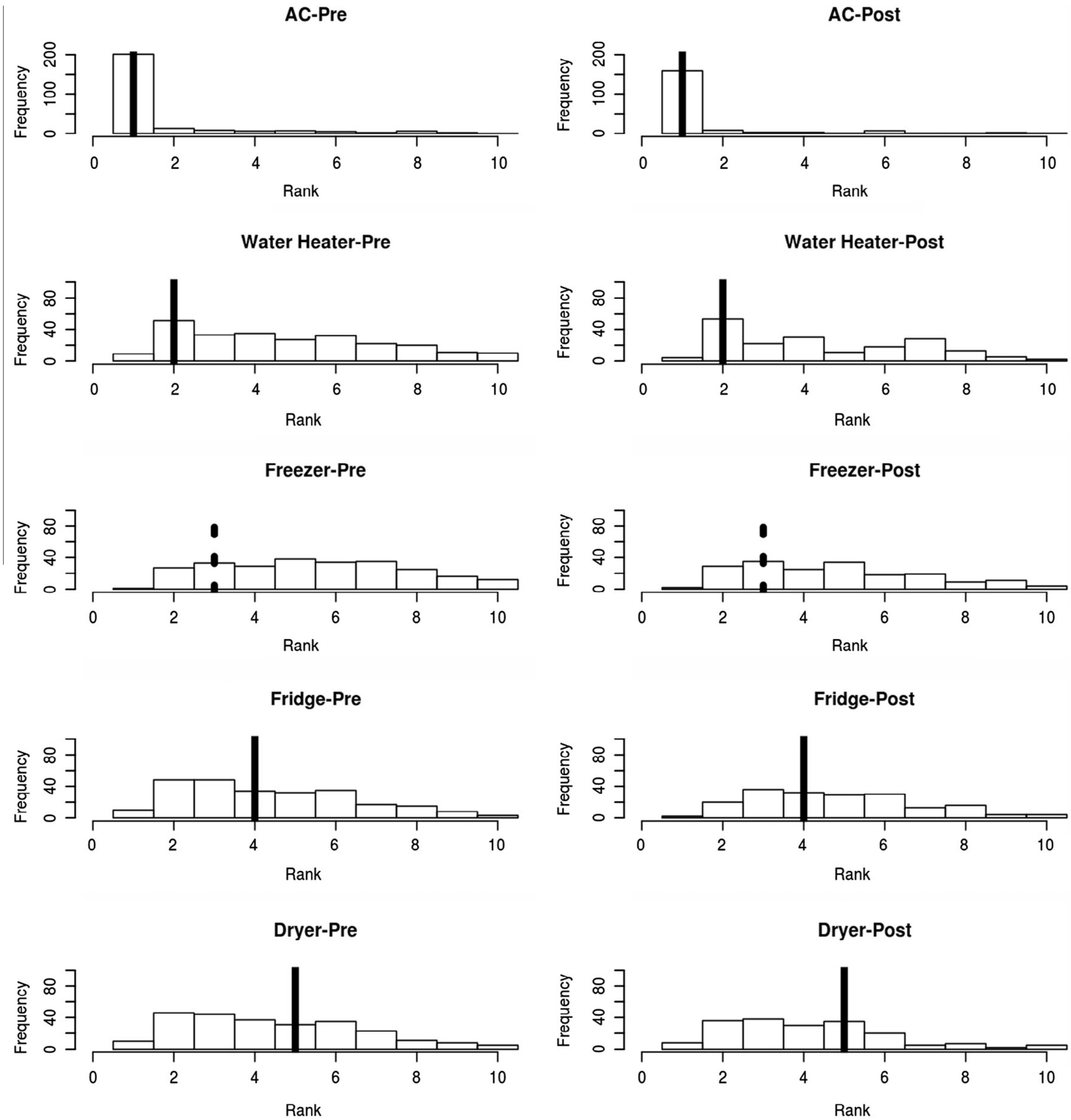

Fig. 4. Histogram of kW h use rankings for each appliance in order of true ranking. Lines indicate true rank, with dashed lines indicating statistically significant pre-post differences.

been unable to focus their attention when provided with a large tabular display that provides information on each appliance. This type of 'information overload' was also discussed in the open-ended comments and has been found elsewhere, especially with respect to the well-established literature on working-memory [47,48], where people have the capacity to consider only three or four 'chunks' of information at one time. Participants learned more about how much power the appliances used when provided with kW h feedback, but not dollars. The ineffectiveness of feedback in dollars on learning energy use indicates that participants could not translate dollars to kW h. These two findings, information overload and inability to translate units, are important because people overwhelmingly prefer bill-to-date (in dollars) and appliance-specific feedback, but we find no evidence of their effectiveness in learning how much energy their appliances use.

Similar to learning how much energy their appliances used, participants were only able to learn how much appliances would cost in a month when provided feedback in dollars. One surprising result, where participants became more accurate in their knowledge of how much power the microwave used but less accurate in their knowledge of how much it costs per month, may provide a window into the learning process. These participants may have based their monthly cost estimates on their power ranking estimates, failing to take into account how often each appliance is used. This explains the opposing effects for the microwave, as learning that it uses more power would also lead one to overestimate kW h usage and subsequently how much it costs per month. One approach to this confusion would be to provide both 'speedometer' and 'odometer'-type feedback information, teaching the distinction between the power an appliance uses and the energy it consumes. However, our findings with respect to information overload would suggest that people would fare better from just receiving projected monthly costs for their appliances, as they have difficulty extrapolating from current cost and energy use to monthly cost. Study One found that projected monthly costs are a feature that participants find desirable, but it is not at the top of their list (it ranked fourth below bill-to-date, appliance-specific feedback, and daily projections).

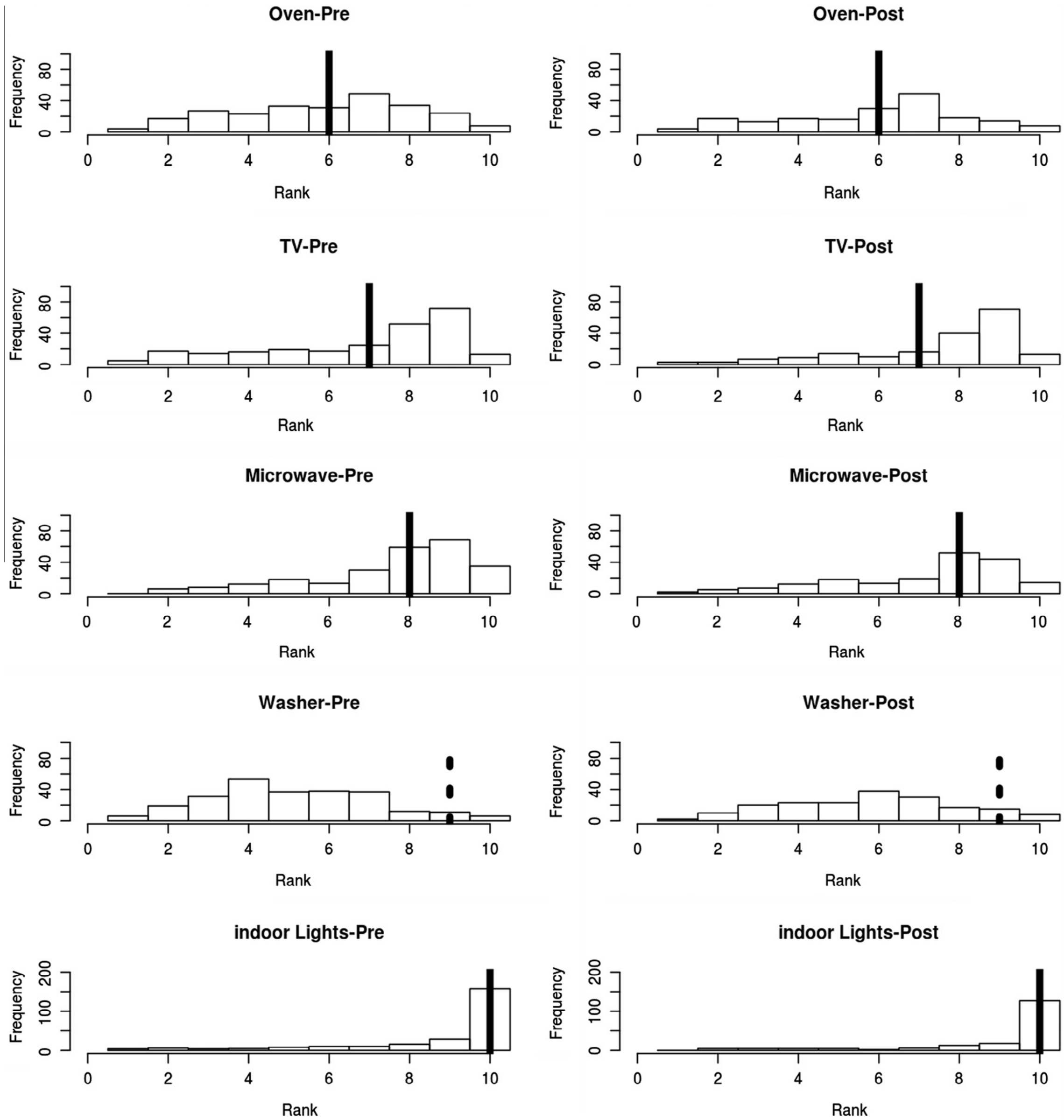

Fig. 5. Histogram of monthly cost rankings for each appliance in order of true ranking. Lines indicate true rank, with dashed lines indicating statistically significant pre-post differences.

Table 3
Hierarchical linear model of the accuracy of appliance rankings by energy use for each condition, with varying intercepts for appliance and participant. Main effects for the presence of kW h information (kW h), appliance-specific feedback (AS), cost ($), and the passive learning condition (passive) are shown in the first, second, third, and fourth rows.

Condition	Improvement	*t*-Statistic	*p*-Value
kW h	0.12	0.75	.30
AS	−0.10	−0.73	.31
$	−0.18	−1.10	.22
Passive	0.65	2.50	.02

Table 4
Hierarchical linear model of the accuracy of appliance rankings by monthly bill contribution for each condition, with varying intercepts for appliance and participant. Main effects for the presence of kW h information (kW h), appliance-specific feedback (AS), cost ($), and the passive learning condition (passive) are shown in the first, second, third, and fourth rows.

Condition	Improvement	*t*-Statistic	*p*-Value
kW h	−0.01	−0.05	.40
AS	−0.06	−0.34	.38
$	0.18	0.84	.28
Passive	0.62	2.60	.01

Perhaps the most surprising finding from the experiment was the consistent success of the passive learning condition. This condition provided participants with the answers to the knowledge test, which they were able to remember and subsequently use to correctly answer questions. This option is attractive, as it is cheaper and logistically easier to give people a flyer with information about appliance-specific energy consumption, monthly cost, and how people should adjust their behavior to save energy, rather than set up a home area network. This solution may not make consumers more knowledgeable about their own electricity consumption, but may induce as much or more behavioral change as an IHD, using a simpler method. However, the external validity of this

Table 5
Identification of the correct unit of energy (kW h) by condition. t-Values compare proportion switched to correct versus incorrect for each condition are based on posterior simulations from a beta (0,0) prior.

Condition	Correct	No change	Incorrect	t-Value
kW h and $ by appliance	5	19	2	0.47
kW h and $ aggregate	7	16	3	0.55
$ by appliance	3	26	2	0.21
kW h by appliance	10	9	4	0.67
Aggregate $ only	3	24	1	0.44
Aggregate kW h only	11	10	2	1.10
Passive learning	9	25	1	1.60
Total	48	129	15	13.00

Table 6
Identification of how energy is calculated by condition. t-Values compare proportion switched to correct versus incorrect for each condition are based on posterior simulations from a beta (0,0) prior.

Condition	Correct	No change	Incorrect	t-Value
kW h and $ by appliance	6	17	3	0.42
kW h and $ aggregate	4	18	4	0.00
$ by appliance	3	21	7	−0.64
kW h by appliance	5	17	1	0.63
Aggregate $ only	3	23	2	0.19
Aggregate kW h only	3	17	3	0.00
Passive learning	10	23	2	1.40
Total	34	136	22	4.80

Table 7
Median estimates of the cost of a kW h before and after treatment. p-Values are from exact Wilcoxon paired rank test on deviations from the true cost ($0.13 per kW h). Z values assume normality. N_t is the number of non-tied (non-zero) deviation scores.

Condition	Median deviation			
	Pre	Post	$Z(N_t)$	p-Value
kW h and $ by appliance	0.57	0.01	1.9 (13)	.06
kW h and $ aggregate	0.26	0.08	1.1 (16)	.27
$ by appliance	0.47	0.17	1.6 (12)	.13
kW h by appliance	0.17	0.12	0.1 (6)	1.00
Aggregate $ only	0.17	0.15	1.8 (11)	.08
Aggregate kW h only	0.87	0.27	0.1 (11)	.97
Passive learning	0.45	0.00	4.5 (33)	.01
Total	0.37	0.07	4.9 (102)	.01

finding is questionable for two reasons. First, participants in the passive learning condition likely did not develop the necessary working knowledge of their appliances to extrapolate to new appliances, and thus any associated feelings of mastery and control over one's environment that might result from a deeper level of comprehension will not emerge. Second, people routinely receive flyers and inserts, which they dispose of without viewing, so this method of providing information may not be useful in the real world because there is saturation in this medium.

4. General discussion

Successfully implementing any new smart grid technology requires consumer understanding and engagement with that tool. While a range of studies have examined the effectiveness of different IHDs among individual populations, less work has been conducted on systematically examining the specific aspects of each IHD that may influence consumer understanding or behavior. We feel that this work is especially important if public regulatory bodies will otherwise be making policy about these technologies based on hard-to-interpret pilot studies or intuition about what might work best for customers.

The studies presented here provide an approach for systematically testing whether consumers learn from specific types of feedback information. While we conducted this work in the context of IHDs, this kind of systematic testing is extremely useful for predicting how consumers will learn from any new technologies that provide electricity use feedback information. Moreover, this approach provides a scientific basis for selecting the kinds of technologies to provide to consumers on a larger-scale.

Our survey and experiment find that consumers report preferences for certain types of information but, in fact, learn better from other types of information. Specifically, consumers showed a strong preference for more detailed appliance-specific feedback in dollar units. However, our results suggest that consumers need to be provided with only an aggregate summary of their current kW h usage to better learn the relative electricity consumption of their appliances. Moreover, providing additional forms of information, for example, providing feedback in both $ units and kW h units, does not help with learning. It may even be the case that providing basic well-designed information may be more effective than learning through interaction with an IHD. While this work illustrates that consumer preferences cannot serve as a substitute for systematic experimental testing, we believe that they provide an important guideline identifying the kinds of information type that should be tested. An understanding of both consumer preference and consumer performance is key to a big picture understanding of consumer needs.

The results from our experimental computer-based IHD simulation show that consumers are already familiar with the relative electricity use of those appliances at the two extremes—the air conditioning and the lights. This may be due to their frequency of use of these appliances. Consumers appear to best incorporate feedback information for these appliances and for the other appliances that they have firm pre-existing beliefs about, regardless of the initial accuracy of their beliefs and regardless of type of feedback they are given. For appliances that are used less frequently, confusion may arise between the electricity required to power the appliance and the amount of energy used by the appliance. It may be that consumers are more willing to learn about appliances they have already given thought to and it may be that there are more concrete mental models about how some appliances work than others.

Based on these two studies, we recommend that in-home displays provide feedback information in those units that are consistent with the desired knowledge goal. Even if consumers report a preference for dollar-unit feedback, kW h feedback will be more effective for learning about relative electricity usage. Due to initial evidence that participants are unable to incorporate frequency of use into predictions about cost, we would suggest providing projected monthly cost by appliance if understanding of relative cost were a knowledge goal. We would caution against the provision of excessive information in an attempt to cover too many knowledge areas. Furthermore, we also caution against selecting specific attributes or in-home displays solely on the strength of these initial findings. A more low-tech and cost-effective form of information provision, such as personalized information directly on a customer's bill (instead of an insert) in an easy-to-understand format, may be just as, if not more effective than an IHD.

There are several additional factors to consider when examining consumer learning from the kind of experimental tool that we have developed. Consumers' pre-existing familiarity with their appliances, their exposure to energy conservation campaigns, and even which appliances are heuristically linked together in their minds may influence their ability to incorporate information. Future studies will be necessary to explore these potential influences on learning. Rigorous and systematic testing of individual information feedback is the only means of identifying the best type of feedback to provide to consumers to achieve increased understanding of electricity usage. Our approach provides a feasible way to do this without the logistical and resource challenges associated with a trial in the field.

Acknowledgements

We thank Jay Apt and Baruch Fischhoff for their thoughtful comments and support, and Ashley Terwilliger for programming the simulation.

This material is based upon work supported by the Department of Energy under Award Numbers DE-OE0000300 and DE-OE0000204. Disclaimer: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights, Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

This work was supported in part by the center for Climate and Energy Decision Making (SES-0949710), through a cooperative agreement between the National Science Foundation and Carnegie Mellon University.

All materials and data, including completely reproducible statistical analyses in Sweave, can be obtained at http://hdl.handle.net/1902.1/19153. Open lab notebook can be obtained at http://openwetware.org/wiki/IHD_Simulation.

References

[1] Kempton W, Layne LL. The consumer's energy analysis environment. Energy Policy 1994;22(10):857–66.
[2] Darby S. Smart metering: what potential for householder engagement? Build Res Inform 2010;38(5):442–57.
[3] Keil FC. Folkscience: coarse interpretations of a complex reality. Trends Cogn Sci 2003;7(8):368–73.
[4] Kempton W, Montgomery L. Folk quantification of energy. Energy 1982;7(10):817–27.
[5] Kempton W. Two theories of home heat control. Cogn Sci 1986;10(1):75–90.
[6] Rozenblit L, Keil F. The misunderstood limits of folk science: an illusion of explanatory depth. Cogn Sci 2002;26(5):521–62.
[7] Wood G, Newborough M. Dynamic energy-consumption indicators for domestic appliances: environment, behaviour and design. Energy Build 2003;35(8):821–41.
[8] Krishnamurti T, Schwartz D, Davis A, Fischhoff B, Bruine de Bruin W, Lave L, et al. Preparing for smart grid technologies: a behavioral decision research approach to understanding consumer expectations about smart meters. Energy Policy 2012;41:790–7.
[9] Seligman C, Darley JM, Becker LJ. Behavioral approaches to residential energy conservation. Energy Build 1978;1(3):325–37.
[10] Seligman C, Becker LS, Darley JM. Encouraging residential energy conservation through feedback. Adv Environ Psychol 1981;3:93–113.
[11] Dobson JK, Griffin JDA. Conservation effect of immediate electricity cost. Feedback on residential consumption behavior. In: Proceedings of the 7th ACEEE summer study on energy efficiency in buildings; 1992. p. 10.33–35.
[12] Davis A, Krishnamurti T, Fischhoff B, Bruine de Bruin W. Setting a standard for electricity field studies. Unpublished manuscript; 2012.
[13] McClelland L, Cook SW. Energy conservation effects of continuous in-home feedback in all-electric homes. J Environ Syst 1979;9(2):169–73.
[14] Yun TJ. Investigating the impact of a minimalist in-home energy consumption display. In: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems. ACM; 2009. p. 4417–22.
[15] Abrahamse W, Steg L, Vlek C, Rothengatter T. A review of intervention studies aimed at household energy conservation. J Environ Psychol 2005;25(3):273–91.
[16] Carroll E, Hatton E, Brown M. Residential energy use behavior change pilot. Technical report, Franklin energy; 2009.
[17] Darby S. The effectiveness of feedback on energy consumption. A Rev DEFRA Lit Meter, Bill Direct Displays 2006;486.
[18] Ehrhardt-Martinez K, Donnelly KA, Laitner S. Advanced metering initiatives and residential feedback programs: a meta-review for household electricity-saving opportunities. Technical report, American Council for an Energy-Efficient Economy; 2010.
[19] Faruqui A, Sergici S, Sharif A. The impact of informational feedback on energy consumption – a survey of the experimental evidence. Energy 2010;35(4):1598–608.
[20] Anderson W, White V. Exploring consumer preferences for home energy display functionality. Technical report; 2009.
[21] Wilson TD, Gilbert DT. Affective forecasting. Adv Exp Soc Psychol 2003;35:345–411.
[22] Loewenstein G, Lerner JS. The role of affect in decision making. USA: Oxford University Press; 2003. p. 619–642.
[23] Steg L, Vlek C. Encouraging pro-environmental behaviour: an integrative review and research agenda. J Environ Psychol 2009;29(3):309–17.
[24] Wilhite H, Ling R. Measured energy savings from a more informative energy bill. Energy Build 1995;22(2):145–55.
[25] Lutzenhiser Loren. Social and behavioral aspects of energy use. Ann Rev Energy Environ 1993;18(1):247–89.
[26] Vassileva I, Odlare M, Wallin F, Dahlquist E. The impact of consumers' feedback preferences on domestic electricity consumption. Appl Energy 2012:575–82.
[27] Hargreaves T, Nye M, Burgess J. Making energy visible: a qualitative field study of how householders interact with feedback from smart energy monitors. Energy Policy 2010;38(10):6111–9.
[28] Langevin J, Gurian PL, Wen J. Reducing energy consumption in low income public housing: interviewing residents about energy behaviors. Appl Energy 2012.
[29] Egan C. Graphical displays and comparative energy information: what do people understand and prefer. In: Summer study of the European council for an energy efficient economy; 1999. p. 2–12.
[30] Wilhite HA, Hoivik A, Olsen JG. Advances in the use of consumption feedback information in energy billing: the experiences of a Norwegian energy utility. In: Proceedings, European council for an energy-efficient economy; 1999.
[31] Roberts S, Humphries H, Hyldon V. Consumer preferences for improving energy consumption feedback. Technical report, Ofgem, Centre for Sustainable Energy; 2004.
[32] Wood G, Newborough M. Energy-use information transfer for intelligent homes: enabling energy conservation with central and local displays. Energy Build 2007;39(4):495–503.
[33] Fischer C. Feedback on household electricity consumption: a tool for saving energy? Energy Efficiency 2008;1(1):79–104.
[34] Roberts S, Baker W. Towards effective energy information. improving consumer feedback on energy consumption. Technical report, Ofgem, Centre for Sustainable Energy; 2003.
[35] Neenan B, Robinson J, Boisvert RN. Residential electricity use feedback: a research synthesis and economic framework. Technical report, Electric Power Research Institute; 2009.
[36] Karjalainen S. Consumer preferences for feedback on household electricity consumption. Energy Build 2011;43(2):458–67.
[37] Ueno T, Sano F, Saeki O, Tsuji K. Effectiveness of an energy-consumption information system on energy savings in residential houses based on monitored data. Appl Energy 2006;83(2):166–83.
[38] Allen D, Janda K. The effects of household characteristics and energy use consciousness on the effectiveness of real-time energy use feedback: a pilot study. In: Proceedings of the ACEEE summer study on energy efficiency in buildings; 2006. p. 7.1–12.
[39] Paetz AG, Becker B, Fichtner W, Schmeck H. Shifting electricity demand with smart home technologies–an experimental study on user acceptance. In: 30th USAEE/IAEE North American conference online proceedings; 2011.
[40] Paolacci G, Chandler J, Ipeirotis PG. Running experiments on Amazon Mechanical Turk. Judgment Decis Mak 2010;5(5):411–9.
[41] Horton JJ, Rand DG, Zeckhauser RJ. The online laboratory: conducting experiments in a real labor market. Exp Econ 2011;14(3):399–425.
[42] Mason W, Watts DJ. Financial incentives and the performance of crowds. ACM SIGKDD Explor Newslett 2010;11(2):100–8.
[43] Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 2011;6(1):3–5.
[44] DeWaters JE, Powers SE. Energy literacy of secondary students in new York state (USA): a measure of knowledge, affect, and behavior. Energy Policy 2011;39(3):1699–710.
[45] Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. New York: Cambridge University Press; 2007.
[46] Gelman A, Su YS, Yajima M, Hill J, Pittau MG, Kerman J, et al. arm: Data analysis using regression and multilevel/hierarchical models. R package version 2010:1–3.
[47] Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav Brain Sci 2001;24(1):87–114.
[48] Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 1956;63(2):81–97.