INTERNATIONAL REAL ESTATE REVIEW

2020 Vol. 23 No. 1: pp. 107 – 136

An Experimental Study of the Effect of Energy Label Design on the Correct Evaluation of Buildings' Energy Performance and Promotion of Energy Saving

Mieko Fujisawa*
Kanazawa University, Kakuma-machi, Kanazawa, Ishikawa, 920-1192, Japan. E-mail: Fujisawa@staff.kanazawa-u.ac.jp

Kazuhisa Takemura
Waseda University, 1-104 Totsuka-machi, Sinjyuku-ku, Tokyo, 169-8050, Japan. E-mail: kazupsy@waseda.jp

Yukihiko Funaki
Waseda University, 1-6-1 Nish-iwaseda, Sinjyuku-ku, Tokyo, 169-8050, Japan. E-mail: funaki@waseda.jp

Nobuyuki Uto
Hokuriku University, 1-1 Taiyogaoka, Kanazawa, Ishikawa, 920-1180, Japan. E-mail: nobuyuki.uto@gmail.com

Ryo Takahashi
Waseda University, 1-104 Totsuka-machi, Sinjyuku-ku, Tokyo, 169-8050, Japan. E-mail: inter.takahashi@gmail.com

Many countries have implemented an energy labeling system that displays the energy savings performance of buildings. However, previous studies have reported mixed findings on the effectiveness of this system.

This study is undertaken in Japan to assess whether energy labels are effective and validate the significance of the label design. We use an online survey to determine the effects of energy labels on target consumers. We choose two energy label designs: one with a stairs rating form design and the second with a rating scale of the energy efficiency. Adopting a within-subjects method, each respondent is

* Corresponding author

shown both designs and asked to evaluate the energy-saving level based on the labels. Cross-tabulation confirms the positive effects of labeling on energy-savings levels, which are amplified by increasing the level of the reference point. The highest insulation performance level in Japan complies with a law enacted back in 1993; the results show that greater effects can be realized by increasing the level of the reference point.

A logistic regression analysis shows that energy labels with the stairs rating have a negative influence, while labels with a rating scale are preferred based on ease of understanding by the respondents. This suggests that energy labels that offer a rating scale can contribute to facilitating energy savings, although there are problems with the comprehension of the gradation colors which need to be further examined. These results can act as a guideline towards optimal designs of the upcoming task of modifying the energy labeling system in Japan.

Keywords

Energy Label, Home Insulation Performance, Real Estate Advertising, Framing Effect, Priming Effect, Reference Point

1. Introduction

In the 2016 Paris Agreement, Japan pledged to the international community that by 2030, Japan would reduce its carbon dioxide (CO_2) emissions by 26% of its 2013 levels. The industrial sector, where measures are progressing, accounts for 7% of the CO_2 emissions in Japan whereas the household sector accounts for 39%. Hence, the Japanese government, in anticipation of the development of energy-saving and zero-energy housing[1], has been focusing on heat insulation performance to realize this drastic reduction in household energy use (e.g., "Plan for Global Warming Countermeasures Cabinet decision on May 13, 2016"[2]).

The "Act on the Improvement of Energy Consumption Performance of Buildings"[3] (passed on July 8, 2015) for reducing CO_2 emission employs the

[1] White Paper on Land, Infrastructure, Transport and Tourism in JAPAN 2019, Chapter 8 Creating and Preserving a Beautiful and Healthy Environment, Retrieved May 20, 2020 from: https://www.mlit.go.jp/common/001325170.pdf

[2] Retrieved May 6, 2020 from: https://www.env.go.jp/en/headline/2238.html

[3] Retrieved May 6, 2020 from: https://www.mlit.go.jp/common/001134876.pdf and https://climate-laws.org/cclow/geographies/japan/laws/act-on-the-improvement-of-energy-consumption-performance-of-buildings

"Energy Efficiency Performance Labeling System" to display the energy consumption performance of buildings, including that of houses. In other words, the government aims to show the energy consumption of housing with an energy-savings performance indication system. This system will encourage consumers to choose high-performing housing insulation (energy-conserving housing) based on evaluation through the information on the energy label. Thus, the energy label can provide information about the energy efficiency of the houses, and the information can assist consumer choices during purchases. Furthermore, the government believes that reimbursing the cost of improvements in thermal insulation, linked to the market price, will be a further incentive to reduce residential energy consumption.

Japanese real estate advertisements will soon mandate the display of an energy label. However, before they are mandated, more discussion is needed on energy label design in Japan, including additional research on how the labels influence the decision-making process of the Japanese in housing selection. Questions remain regarding the feasibility of the government to reach their goals by promoting consumer choice of energy-saving housing through an energy label that indicates the energy consumption performance of a house. Measuring the effects of energy label use is important for achieving the target CO_2 reduction. Thus, we must determine whether it is possible to promote the selection of energy-saving housing through the use of energy labels. To do so, this study adopts a two-step verification process: (1) clarifying the design conditions to correctly understand the information on energy labels, and (2) identifying whether energy labels change consumer choices. If the ingenuity of energy label design influences housing selection, Japan can reduce CO_2 emissions without a fiscal stimulus, such as government subsidies. On the contrary, if labels do not influence consumer selection, then the CO_2 reduction plan should be reconsidered.

This study measures the relationship between the evaluation and design of energy labels toward the selection of energy-saving housing. The premise of this study is that if energy-saving labels are well designed, they will help consumers to understand the evaluations more accurately from the label information and guide them to make better choices. The hypothesis in this study is that the optimum design of energy labels lead to the selection of housing with higher energy-saving levels. An experimental survey and analysis are conducted to test this hypothesis. We conduct a logistic regression analysis with experimental data to explore the determinants and label design factors that influence the correct reading of energy consumption information by consumers.

This paper is organized as follows: Section 2 summarizes and compares the components of this study with those of previous studies. Section 3 provides the experimental data and explains the analytical method. Section 4 discusses the results of the cross-tabulation and logistic regression analyses. Section 5 concludes the study and discusses the policy implications for effective designing and displaying of energy labels.

2. Previous Studies vs. This Study

We first review the literature on energy labels and then summarize the findings based on behavioral economics. The research gap is then presented, which will be addressed in this study.

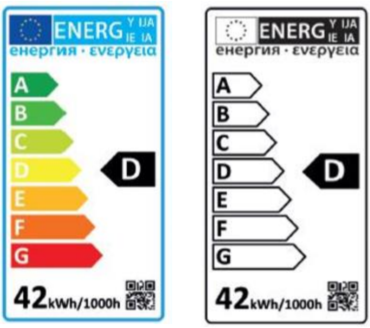
2.1 Literature

Consumers generally lack knowledge of products or services compared to the companies that produce them. This is widely recognized as information asymmetry which causes market failure. To prevent this issue, information and labels are used to remove information asymmetry (Bull, 2012).

Information on the energy-saving performance of houses falls under a policy tool called energy performance certification (EPC). In the European Union (EU), disclosing the EPC when conducting housing transactions has been mandatory since 2013 (Murphy, 2014). Fuerst and McAllister (2011) confirm the importance of EPC and then its effect before it is made mandatory. Murphy (2014) and Fuerst et al. (2015) state that there is a positive relationship between high energy-saving performance and housing price after EPC became mandatory.

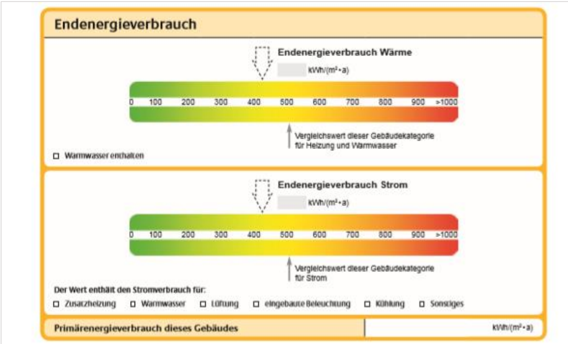
The energy label provides a simple summary of the EPC. Some of the energy label designs that are being used include for example, the stairs rating form design (see Figure 1) and a rating scale of the energy efficiency (see Figure 2) in the EU. The two labels are named based on their design. Many EU countries, such as the United Kingdom and France, have adopted labels with the stairs rating.

Figure 1 Examples of Stairs Rating (Color and Black & White)



Source: European Commission (2020). Study on the impact on consumer understanding and purchase decisions of energy labels for lighting products Final Report, p 13. https://ec.europa.eu/energy/sites/ener/files/final_report_energy_labels_-_lighting_products.pdf (Accessed March 18, 2020)

Figure 2 Example of Label with Rating Scale



***Source*:** Enviro German Energy Agency. https://www.zukunft-haus.info/fileadmin/Zukunfthaus/Bilder/Beratung_Planung/Energieausweis/Formular_EA_WG_2014.pdf (Accessed May 17, 2020)

In countries where energy labeling is mandatory, consistency is found in labels that are also used for goods such as household appliances, thus simplifying the calculation of expected energy usage (Bull, 2012). Stadelmann and Schubert (2018) and Blach et al. (2019) conclude that energy labels on appliances have a positive effect. Energy labels have been introduced as a guideline for energy conservation and therefore to encourage energy saving behaviors not only in the EU but also other countries (Wiel and McMahon, 2005). In the United States, the Energy Star program provides guidelines and standardization methods for energy labels. Murray and Mills (2011) and Ward et al. (2011) analyze the effectiveness of the Energy Star program. In addition, Walls et al. (2017) examine the effect of the Energy Star program on American housing prices and find that certification under this program results in higher house prices in certain markets. Similarly, Brounen and Kok (2011) show that there is a positive relationship between housing price and energy labels in the EU.

Studies have also shown that consumer choices based on energy labels might not always be optimal. Waechter et al. (2015) show that consumers choose larger refrigerators instead of high-efficiency ones due to misinterpretation of the energy label, in which they base their decision on the energy efficiency class instead of the annual electricity consumption. Andor et al. (2017) examine whether consumers would purchase energy efficient consumables if the operating cost of appliances is shown on the energy label. They find that consumers pay higher initial costs for energy-efficient appliances and make their choices based on the operating cost.

Prior to 2010, the energy label used in the EU was a color scale with a letter sequencing from A to G. After 2010, the label was revised to a pictogram instead of using words so that the products and devices can be marketed in different EU countries without translation. Moreover, the A rating has been further expanded to A+ to A+++ to include higher efficiency ratings. However, Heinzle and Wustenhagen (2012) examine the effect of using letter ratings by comparing the old and new labels, but could not confirm any significant effect. Another study by Waechter et al. (2016) suggests that the expansion of the A rating might have increased the importance of energy efficiency, but consumers do not choose energy-efficient devices more after this change in the label.

Bjerregaard and Møller (2019) examine consumer behavior before and after the mandatory implementation of energy labels in Denmark and conclude that sales of highly efficient cold appliances have increased, which shows response to civic responsibility. However, Thonipara et al. (2019) state that the effects of energy policies are not the same in all of the EU countries where it is mandatory to have energy labels in advertisements, and changes in consumer behavior greatly depend on the introduction of a carbon tax instead.

There is currently no study in Japan that supports the implementation of an energy labeling system in the near future, but many studies have been conducted on energy labels in other countries where the system has already been implemented. Energy labels can act as a guideline for energy conservation, and have energy-saving effects premised on what Thaler and Sunstein (2008) and the Behavioural Insights Team (2011) label as a nudge. Nudging is a psychological means of influencing choices based on arranging choices so that they are presented in a certain way that elicits a certain response or behavior. In marketing, nudging strategies eventually facilitate changes in consumer behavior. Ölander and Thøgersen (2014) consider previous research on the effects of informing (providing information to consumers) versus nudging on consumers, and find that the use of both is important.

Energy labelling could also be affected by how the information is framed. Plevin et al. (1988) show the presence of an information framing effect on the choice behavior of consumers; that is, choices change depending on how information is conveyed. Van de Velde et al. (2010) conclude that how a message is framed has impacts on consumer behavior, and find that emphasizing a positive frame has a stronger impact than using a negative frame. In addition, Mandel and Johnson (2002) conclude that priming can affect decisions based on product selection behavior of consumers after the background images and colors of a website are changed. Wang (2011) also emphasizes the importance of the mutual use of framing and priming effects.

In terms of the format of energy labels, a study by Waechter et al. (2016) which focuses on letters, signs, and colors to determine the factors that affect consumer choices finds that a shorter scale is optimum and therefore the label format is important for motivating the selection of products that are energy efficient. Andor et al. (2017) examine purchases of energy-efficient products with labels that show additional details on annual operating cost and find that consumers in Germany were willing to pay EUR 30 more for energy-efficient refrigerators as a result.

According to Kahneman et al. (1982), individuals show cognitive bias in behavioral economics. They argue that it is possible to change consumer choices through this bias because of the anchoring effect. When consumers start to look for new products to purchase, first and foremost, they make judgement based on their own biases. The strongest bias is the anchoring effect which serves as the reference for judgement. Energy labels can affect this bias, when consumers select a product based on the energy label information or by using the information as a reference point (Northcaft and Neale, 1987, Bucchianeri and Minson, 2013).

However, the actions of consumers related to saving energy or protecting the environment are known to evolve over time. Gillingham et al. (2016) state that the behavior of consumers is reversed, as shown by the rebound effect, after their initial actions to save energy or protect the environment. However, a properly designed energy labels can encourage consumers to select an energy-saving house which will solve the rebound effect, because if consumers select an energy-saving house which consumes less energy at the outset, the effect will continue without being directly associated with consumer behavior. This is the significance of our study.

2.2 Energy Labels and Cognitive Bias

In this study, we examine: (1) the design of energy labels and (2) the possibility of encouraging consumers to choose high-performing housing insulation (energy-conserving housing) through cognitive bias. The optimum design of an energy label is determined by examining the differences in color, dimensions (width) and wording of the label based on priming and framing. The probability of convincing consumers to opt for energy-conserving housing is experimentally determined by the use of reference points, which is based on the prospect theory. This theory states that individuals measure value in relation to a reference point; that is, a standard that is used for comparison or evaluative purposes (Kahneman and Tversky, 1979). Therefore, it is assumed that human values can be changed through cognitive bias by changing the level of the reference point. First, assuming that the anchoring effect is based on the reference point, we determine the probability of changing consumer behavior. Next, we use a logistic regression model to analyze the responses of consumers to changes in the level of the reference point.

This study experimentally compares the use of a rating scale on energy labels with a stairs rating, which are both commonly used in the EU. The rating scale has become increasingly common after it was introduced as an energy label sample by an external partner institute of the Japanese government[4]. The purpose of this experiment is to determine the impact of the energy label design and the selection behavior of consumers by using both the rating scale and the stairs rating. Specifically, this study creates an energy label design that features two elements in the original energy labels. One element is the standards for the primary energy consumption of a building or housing, and the other shows the rate of reduction from the original amount of consumption. We consider the former element as the reference point and create an experimental plan that captures the reactions of consumers to the changes in the level of the reference point based on their evaluation of the label.

No similar studies have been done in the literature that compare the design of energy labels and evaluates the use of the stairs rating vs. the rating scale on energy labels and thus, these are the significant contributions of this study.

3. Method

The design of the experiment and the study, as well as results of an online experiment will be discussed in the following section.

3.1 Experimental Design

The experimental frame of this study and details of an online survey are discussed in the following.

3.1.1 Experimental Frame

The experiment design is based on several considerations, including the effects of priming and framing, and the within-subjects and order bias methods.

1) Priming effect: The effect of energy labels is examined by focusing on the different colors and dimensions to create three different labels: “black & white” with the “same width”; “black & white” with a “different width”; or “colored” with a “different width”.
2) Framing effect: The effects of the reference point are labeled as “average level”, “ideal level”, and “legal standard level” and used because they are the standards to denote primary energy consumption.
3) Within-subjects method: The within-subjects design is one where the respondents are provided with every condition, and in this study, means that each subject responds to questions for both types of energy labels. The within-subjects design also makes it possible to compare and analyze

[4] Retrieved May 6, 2020 from: https://www.hyoukakyoukai.or.jp/bels/bels.html (in Japanese).

the findings of both labels, with fewer errors from their individual effects (Charness et al. 2012).

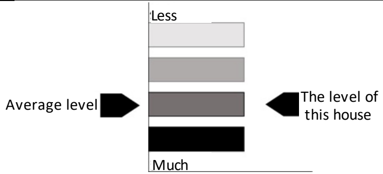
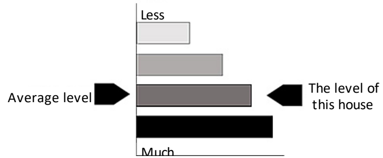
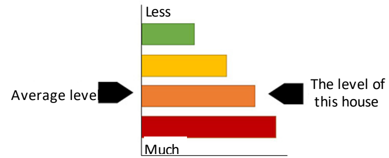
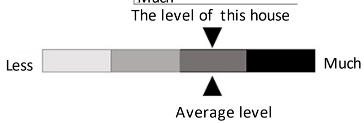
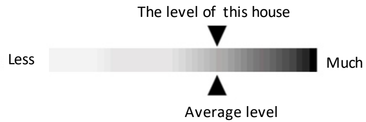
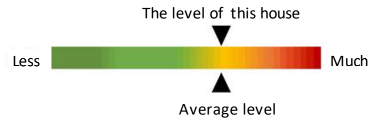
4) Order bias: To obtain the responses to the design of the energy labels and eliminate the bias of the ordering of the questions, two types of surveys are used based on the ordering of the survey questions: one that starts with the questions related to the label with the stairs rating (hereinafter 'Treatment A') and the second related to the label with the rating scale (hereinafter 'Treatment B'). Therefore, the nested energy labels avoided order bias (Serenko and Bontis 2013).

Based on the above, we design and examine the labels which offer six designs in total (see Appendix A for the design type used for each survey question). Table 1 lists the elements of the six designs, and Table 2 shows the specific types of the six designs.

Table 1 Six Designs

Type	Type	Color		Width	
		Black & white	Color	Same	Different
Black & white	1	○		○	
Same width	2	○		○	
Black & white	3	○			○
Different width	4	○			○
Color	5		○		○
Different width	6		○		○
Type	Type	Treatment A		Treatment B	
		Stairs Rating Form Design	Rating Scale	Stairs Rating Form Design	Rating Scale
Black & white	1	○			○
Same width	2		○	○	
Black & white	3	○			○
Different width	4		○	○	
Color	5	○			○
Different width	6		○	○	

Table 2 Types of Energy Labels

Design	Black & white Same width	Black & white Different widths	Colored Different widths
Energy efficiency stairs rating form design			
Rating scale of energy efficiency			

Note: The arrows represent the level of energy consumption.

3.1.2 Respondents

The participants in this study consist of those who had purchased a house between 2012 and 2017. They were recruited through a questionnaire administered by an online survey company. We used an online survey company because they agreed to use their own attributes (data) with registered monitors of the company. As stipulated, we received the data only after the information was anonymized, thus addressing any ethical issues.

A pilot paper survey was first conducted with students, and a subsequent investigation was designed and carried out after incorporating their feedback. The pilot survey was presented in a PowerPoint (PPT) presentation format accompanied by a paper questionnaire, and the survey period was August 21 to 25, 2017 and January 16, 2018. The pilot surveys were conducted with student collaborators through individual interviews. We presented the PPT slides that explained about the energy labels and then asked them to complete the questionnaire. The survey took place in the lecture rooms of certain universities, where only the respondents were present.

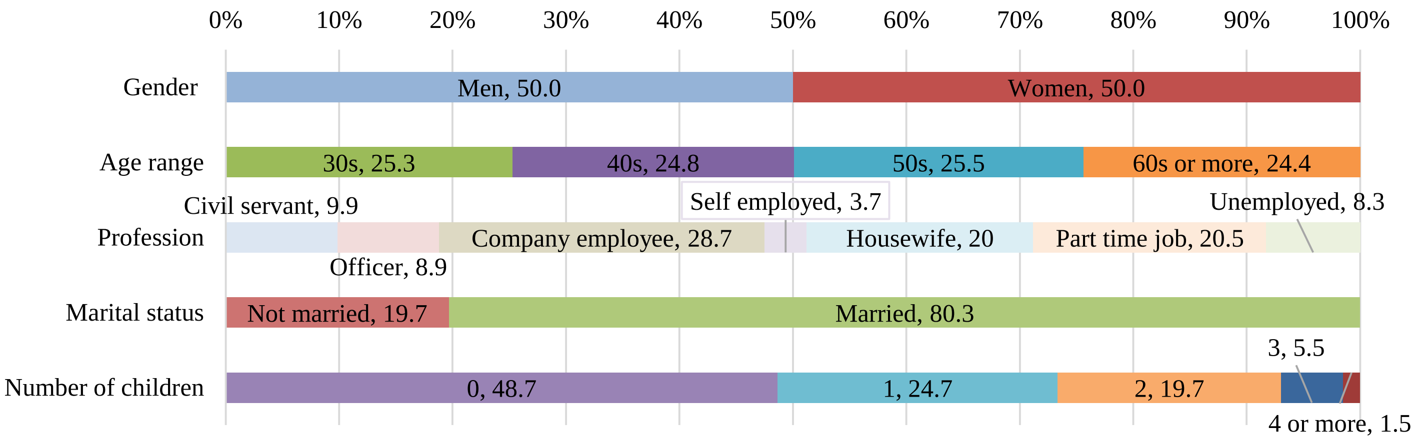
Table 3 shows the outline of the online survey. The actual survey was conducted on March 13, 2018. The number of requests made for monitors was 2668, while the number of valid responses was 1078 (a response rate of 40.4%). There were 18 questions on the survey which revolved around label evaluation, visibility of the label, and environmentally conscious behavior. Attributes such as gender and age were included when the respondents registered as monitors.

Figure 3 presents the demographics of the respondents. There are an equal number of male and female respondents. Also, the age range of the respondents almost parallels the actual demographics in Japan, as we modified the design of the survey to accord with those statistics. By profession, half of the participants are full-time workers, approximately a quarter are part-time workers, and the rest are unemployed. The majority (80%) are married while half of them are families with children. The survey target are consumers who have purchased a house between 2012 and 2017.

Table 3 Survey Outline

Name of survey	Survey on energy label design
Survey method	Web questionnaire survey
Survey agency	INTAGE Inc.
Date of survey	March 13 (Wed.), 2018
Target of questionnaire	Consumers who have purchased a house between 2012 and 2017
Treatment item	Determining the attractiveness of a house through energy labels Selection of easy-to-read energy label Questions about eco-friendly behavior

Figure 3 Demographics of Respondents



3.2 Study Design

We proceeded with the approach of using two designs to determine the effects of the label design and the probability that the designs will motivate the selection of energy-saving options. For this purpose, we use cross-tabulation and logistic regression analyses.

3.2.1 Validating the Effects of Label Design

The data were cross-tabulated to compare the effects of the labels with the stairs rating and a rating scale. In particular, the energy label designs were compared in terms of the effects of priming, framing, and anchoring.

In addition, simple text that indicated the level of energy consumption was printed onto the label to accurately measure the effects. Based on the results of the pilot survey with the students, we used phrases such as “much” and “less” in the questionnaire (see Appendix B for the differences between the pilot and online surveys). One of the reasons for the immense confusion in the pilot Treatment A in terms of the label with the rating scale is that respondents misinterpreted the graphics that denote the level of energy consumption. Another reason is that showing the level of energy consumption on the label influences judgment (Stadelmann and Schubert, 2018; Blach et al., 2019).

(1) Priming Effect
In this study, the priming effect is determined by examining the confusion and the stray rates. The confusion rate is indicated by the number of respondents who answered “yes”, even though the energy consumption of the house is lower than the reference point. Respondents likely have confused the standards as increased versus decreased consumption; they may not have understood the energy usage graphics or the meaning of the standards. Thus, it is assumed in this study that a lower confusion rate means that the label can be readily interpreted with ease.

The stray rate denotes the number of respondents who replied, “I do not know”. The reasons for giving this answer could include “I cannot evaluate the item”, “I do not agree with my own values”, “I do not want to evaluate”, and so on and so forth. However, if intuitive judgment could not be made, this indicates that there is an issue with the label design. Thus, it is assumed that a lower stray rate means that the label can be readily interpreted.

(2) Framing Effect
In this study, the framing effect is determined by using the confusion and stray rates for different words used at the reference point. We examined the words used to express the reference point that act as the benchmark. The level of the reference point was expressed in three ways: the “average level”, “ideal level”, and “legal standard level”; however, “average level” was not used in the

analysis. Cross-tabulation was used to check the confusion and stray rates with the two phrases - ideal and legal standard levels.

(3) Anchoring Effect

In this study, the anchoring effect is determined with the reaction rate, which indicates the possibility of reaching a higher standard of energy conservation. The reaction rate is determined by using two reference points: “increasing the standards” and “increasing the level of the reference points”. In both cases, a higher rate of reaction is assumed to be desirable as the energy label is considered to have the ability to guide users more and conserve more energy.

Increasing the standards means that the standard ratings are increased from four to five, while the energy consumption levels remain the same on the label design. This measures the reaction rate; that is the rate in which respondents change their assessment, and determines the probability of eliciting the compliance of consumers to conserve energy.

The level of the reference point is increased from four to five levels on the label design. This addition of another level is also to determine the probability of eliciting the compliance of consumers to conserve energy.

3.2.2 Determining Probability of Eliciting Compliance

A logistic regression analysis was performed after the cross-tabulation. The analysis focused on an increase in the level of the reference point, and a dependent variable was used with binary variables (0: none, 1: present) to indicate the probability that the respondents are more geared towards energy-saving when the level of the reference point is increased. This indicates the probability that they would be receptive to a home with a higher level of energy efficiency by increasing the level of the reference point. However, the logistic regression analysis strictly explores the determinants and label design factors that influence the correct reading of the energy consumption information by consumers. The logistic model is described below:

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i, \quad (0 < P < 1)$$

where p is the probability of energy-saving tendency as a dependent variable, β is a regression coefficient, and X_i (i=1, …, n) is the i^{th} independent variable. A logistic regression was performed for three models with control variables such as sex, age, and household annual income. The three models consist of Model 1 (those who responded to Treatment A), Model 2 (those who responded to Treatment B), and Model 3 (combined responses to both Treatments A and B). The descriptive statistics are shown in Table 4.

The independent variable "environmental concern" was calculated based on the responses to the related items on the questionnaire, such as attitude and energy-saving behavior. The dummy of Treatment A identifies a respondent who answered Treatment B after Treatment A. Since the Treatment A is conducted by using the within-subject method, each respondent was shown both label designs.

The dummies of the dimensions of the label, that is the dummies of the different widths, and dummy of the color are meant to identify differences in results due to variations in the design as the priming effect.

Table 4 Descriptive Statistics

	Min	Max	Mean	Standard deviation
Gender dummy (0: Female, 1: Male)	0	1	0.48	0.500
Age (Year)	30	69	49.69	11.458
Education higher than college degree dummy (0: No, 1: Yes)	0	1	0.11	0.315
Married dummy (0: No, 1:Yes)	0	1	0.80	0.402
Number of children (Number)	1	5	1.85	0.982
Household yearly income (10,000 Japanese Yen)	50	2,250	654.88	427.338
Unemployment dummy (0:No, 1: Yes)	0	1	0.30	0.458
Environmental interest (Ref point)	0	10	3.22	2.022
Detached house dummy (0: No, 1: Yes)	0	1	0.71	0.455
Cold region dummy (0: No, 1: Yes)	0	1	0.13	0.340
Temperate region dummy (0:No, 1: Yes)	0	1	0.03	0.160
Treatment A respondent dummy (0: No, 1: Yes)	0	1	0.48	0.500
Different width dummy (0: No, 1: Yes)	0	1	0.67	0.470
Color dummy (0: No, 1: Yes)	0	1	0.34	0.474
Good evaluation of stairs rating form design dummy (0: No, 1: Yes)	0	1	0.33	0.471
Response time (Minutes)	2.16	29.30	6.0129	3.65416

4. Results

This section summarizes the cross-tabulation and logistic regression analysis results. We consider the effects of the label design and enhancements in energy labeling systems based on these results.

4.1 Results of Design Effect

First, we calculated the confusion and stray rates of each design and confirmed the readability of the energy labels. Next, cross-tabulation was performed on increases of a standard and in the level of a reference point, and the possibility of motivating consumers to choose higher standards of energy conservation was evaluated.

All of the results of the cross-tabulation were obtained after performing chi-square tests and statistical confirmations.

4.1.1 The Effect of Priming

Table 5 show the results of the cross-tabulation for each design, which include the confusion and stray rates.

The confusion rate is statistically significant for only color and different widths at the 10% level according to the results of the chi-square tests. Although the confusion rate for the label with a rating scale is higher for both color and black & white, which suggests that the label with a rating scale is more likely to be misinterpreted than the label with the stairs rating, this result is not statistically significant. The difference in the confusion rate between the two labels is statistically significant for only color and different widths. As a result, we assume that the label with a rating scale is more difficult to understand than the label with the stairs rating due to the gradation colors.

The stray rates are statistically significant at the 1% level based on the results of the chi-square tests (see Table 5). The stray rate of the label with the stairs rating is higher than that of the label with a rating scale for all cases. This result suggests that consumers find it difficult to understand the label with the stairs rating, and thus the label with the stairs rating has limitations in accurately conveying information for evaluation purposes.

Table 5 Confusion and Stray Rates

	Stairs rating (%)	Rating scale (%)
Confusion rate		
Black & white, Same width	12.0166	13.8122
Black & white, Different width	13.1944	12.3596
Color, Different width	11.3260*	13.5593*
Stray rate		
Black & white, Same width	18.4392***	14.5028***
Black & white, Different width	19.0278***	14.6770***
Color, Different width	18.0249***	13.7006***

Notes: (a) p<0.01(***) indicates significance at the 1% level of the difference between the stairs rating form design and rating scale for each design based on a chi-square test. In the same test, p<0.05(**) and p<0.10(*) indicate 5% and 10% significance, respectively.
(b) The statistical result of the differences among the type of design is obtained at P<0.01 based on a chi-square test.

4.1.2 The Effect of Framing

The effect of the wording of the reference point is summarized for each design with respect to the ideal and legal standard levels, as shown in Table 6.

The confusion rate on wording differences is higher at the ideal level than the legal standard level for both designs, which is a statistically significant result (see (a) and (b) under Table 6). This suggests that consumers feel that they need to comply with legislation. Moreover, similar to the comparison on the design, the confusion rate of the label with a rating scale is high regardless of how the text is expressed (see (c) under Table 6). The stray rate is not statistically significant, thus suggesting no relationship with either the wording used or the design of the energy label.

Table 6 Wording Confusion and Stray Rates

	Ideal level (%)	Legal standard level (%)
Confusion rate		
Stairs rating	8.9981**	6.1224**
Rating scale	11.8738***	7.5139***
Stray rate		
Stairs rating	13.0798	12.8015
Rating scale	13.7291	12.2449

Notes: (a) $p<0.01$(***) indicates significance at the 1% level of the difference between the ideal and legal standard levels for each design based on a chi-square test. In the same test, $p<0.05$(**) and $p<0.10$(*) indicate 5% and 10% significance, respectively.
(b) The statistical result of the difference between the stairs rating form design and rating scale in each design is obtained at $P<0.01$ based on a chi-square test (only for the stray rate).

4.1.3 The Effect of Anchoring

The reaction rate is determined for each label design with respect to increases in the standard and level of the reference point, as shown in Table 7.

It can be observed that the respondents react more strongly to an increase in the level of the reference point than to increases in the standard, regardless of the label design (see (d) under Table 7).

On the one hand, the label with a rating scale receives a higher response rate than the label with the stairs rating in terms of an increase in the standard, but the difference is not statistically significant (see (b) under Table 7). On the other hand, the label with a rating scale receives a higher response rate when the level of the reference point is increased. However, this is statistically significant only in the black & white label and those with a different width (see (b) under Table 7). The chi-square test results of both designs, which combines the designs with all different colors and widths, are statistically significant at the 1% level (see (c) under Table 7). This suggests that the label with a rating scale is more effective than the label with the stairs rating based on the response rates.

Table 7 Reaction rate

	Stairs rating	Rating scale
Addition of standard		
Black & white, Same width	1.1050	1.1050
Black & white, Different width	1.9444	3.0899
Color, Different width	1.6575	2.2599
Increase in level of reference point		
Black & white, Same width	37.2928	42.2652
Black & white, Different width	36.9444*	45.5056*
Color, Different width	38.6740	41.2429

Notes: (a) $p<0.01$(***) indicates significance at the 1% level of the difference between the stairs rating form design and rating scale for each design based on a chi-square test. In the same test, $p<0.05$(**) and $p<0.10$(*) indicate 5% and 10% significance, respectively.

(b) The statistical result of the differences among the design types is not statistically significance based on a chi-square test.

(c) The statistical result of the difference between the addition of standard and increase in level of reference point is obtained at $P<0.01$ based on a chi-square test.

4.2 Probability of Motivating Consumers to Higher Standards of Energy Conservation

Table 8 shows the results of the logistic regression analysis for the three models; Model 1 which comprises the results of Treatment A, Model 2 the results of Treatment B, and Model 3, the results from both Models 1 and 2. The distinction rate is 56.1% for Model 1, 58.2% for Model 2, and 57.8% for Model 3. According to the Hosmer-Lemeshow test, the significance probability of Model 1 is 0.904, Model 2 is 0.865, and Model 3 is 0.943. Since the variable of the dummy for Treatment A in Model 3 is not statistically significant, the order bias is considered to be addressed.

Table 8 Analysis Results

	Model 1			Model 2			Model 3		
	B	Standard error	Exp(B)	B	Standard error	Exp(B)	B	Standard error	Exp(B)
Gender dummy (0:Female; 1: Male)	-0.033	0.198	0.968	0.306	0.198	1.358	0.040	0.128	1.041
Age (Years old)	0.008	0.009	1.008	-0.008	0.009	0.992	-0.004	0.006	0.996
Education higher than college degree dummy(0：No，1：Yes)	-0.092	0.270	0.912	0.004	0.270	1.004	0.007	0.175	1.007
Married dummy(0：No，1：Yes)	0.213	0.253	1.238	0.018	0.254	1.018	0.126	0.165	1.134
Number of children (Number)	0.039	0.104	1.040	-0.078	0.104	0.925	0.060	0.066	1.062
Household yearly income (10,000 Japanese Yen)	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
Unemployment dummy(0：No，1：Yes)	-0.400 *	0.230	0.670	0.038	0.229	1.039	-0.113	0.149	0.893
Environmental interest (Ref point)	0.086 *	0.046	1.090	0.013	0.046	1.013	0.108 **	0.046	1.114
Detached house dummy(0：No，1：Yes)	0.034	0.209	1.034	-0.081	0.210	0.922	0.036	0.132	1.036
Cold region dummy(0：No，1：Yes)	0.046	0.281	1.048	0.513 *	0.296	1.671	0.198	0.180	1.219
Temperate region dummy(0：No，1：Yes)	-0.060	0.525	0.942	-0.368	0.519	0.692	-0.187	0.351	0.830
Treatment A respondent dummy(0：No，1：Yes)	-0.577 ***	0.183	0.561	-0.127	0.183	0.881	-0.249 **	0.118	0.779
Different width dummy(0：No，1：Yes)	0.092	0.226	1.097	-0.140	0.224	0.870	0.074	0.143	1.077
Color dummy(0：No，1：Yes)	0.029	0.225	1.029	0.227	0.224	1.254	0.142	0.145	1.153
Good evaluation for Stairs Rating Form Design dummy(0：No，1：Yes)	-0.101	0.197	0.903	-0.265	0.196	0.767	-0.054	0.127	0.947
Response time(Minutes)	-0.003	0.025	0.997	0.044	0.027	1.045	0.000	0.001	1.000
Treatment A dummy	-	-	-	-	-	-	0.110	0.117	1.117
Constant	-0.324	0.591	0.723	0.626	0.591	1.870	-0.062	0.382	0.940
Model 1 and 2, N＝648 Model 3, N = 1,214	Distinction rate：56.1% Significance probability：0.904			Distinction rate：58.2% Significance probability：0.865			Distinction rate：57.8% Significance probability：0.943		

Note: Significance level: ***: P<0.01, **: P<0.05, *: P<0.1, The results show a difference in the trends between Treatments A and B. In Treatment A, the respondents who are highly concerned about the environment are more likely to choose housing that offer higher energy-saving levels, with an odds ratio of 1.09. However, the respondents who are unemployed or prefer the label with the stairs rating with Treatment A show a negative reaction to energy savings. The respondents who live in the colder regions and answered Treatment B are more likely to be motivated by energy savings, with an odds ratio of 1.671. The other variables are not significant, thus suggesting that there are some differences between the mentality of those who completed Treatments A vs. B.

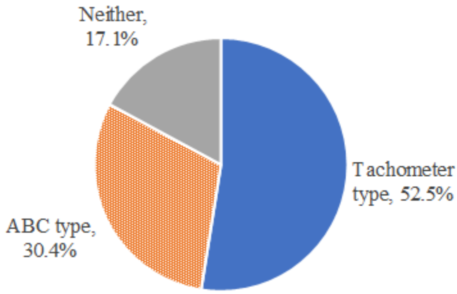
Model 3 shows that the respondents who are highly concerned about the environment and those who answered Treatment A are correlated with the reaction rate. It can be easily shown that the respondents who are concerned about environmental protection issues place more importance on energy saving, which is similar to the results of Model 1. However, despite controlling the experiment by using the within-subjects method and a dummy variable of Treatment A, the responses in Treatment A show that the respondents would not prefer housing that offers a higher energy-saving level. This means that it will be difficult to convince consumers to make choices that would favor a higher energy-saving level with the label that uses the stairs rating. We hypothesize about the some of the reasons for this negative result, including factors that are related to the lack of familiarity of the respondents with the energy labels before answering the survey questions. This is attributed to the priming effect noted in Mandel and Johnson (2002), but needs to be validated through further research.

The dummy variables of color and different dimensions (width) are not statistically significant in all three models. These results suggest that the design differences have no effects on preference for housing that offers a higher energy-saving level.

4.3 Discussion

The results of the cross-tabulation show that the purpose of the rating scale is often misinterpreted. Although the confusion rate is higher than that of the label with the stairs rating, this is only due to issues in the gradation color design according to the statistical analysis. We hypothesize that reducing the confusion rate associated with the rating scale after modifying the label based on the pilot survey would be successful due to the use of text to inform the user of the amount of energy consumed, which illustrates the idea of using “more” or “less” in the label design (see Appendix B). Nevertheless, the respondents still felt that the label with the rating scale is easier to understand, as shown in Figure 4. The label with the rating scale seems to be favored by the respondents despite some of the initial confusion. As this result is controlled for by eliminating the order bias, this means that the label with the rating scale is easier to understand even though the anchoring effect is taken into account which increases the influence of what the respondents have already been exposed to in the first place (Bucchianeri and Minson, 2013).

Figure 4 Label That is Easier to Read



The priming effect of the color and dimensions (width) of the labels cannot be confirmed. The measurements of the influence of the design (dummies of color and dimensions (width)) are not statistically significant. Namely, there is no specific influence of the design color or dimensions (width). However, we hypothesize that simply considering the color and dimensions is meaningless, and the overall label design is the most important factor based on the results shown in Table 5.

We found that the framing effect of the reference point words might possibly change respondent behavior. This is similar to the results in Bjerregaard and Møller (2019) on consumer behavior changes when the related legislation was amended. If the text of the reference point indicates a legal standard value, we anticipated that the respondents might tend to choose housing with higher energy savings as they understand the evaluations more and accept their civic responsibilities to a greater extent.

The reference point has an important meaning. When the level of the reference point is increased, there is the tendency of the respondents to choose housing with higher energy savings. Similar results have been obtained in many studies that have confirmed an anchoring effect (e.g., Tversky and Kahneman, 1974). We found that it was more effective to raise the reference point of the insulation performance level than to add an additional upper rank. Since the highest level of home insulation performance is still determined based on the legal stipulations of 1993 in Japan, the results show that greater effects can be produced by increasing the level of this reference point.

The respondents found that it is difficult to understand the label with the stairs rating. The stray rate of the label with the stairs rating is higher than that of the label with a rating scale and statistically significant, thus showing evidence of the respondent preference for the label with a rating scale. In addition, the logistics regression analysis showed that a label with the stairs rating has a negative effect on the likelihood of persuading users to choose housing with a higher energy savings level. That is, an energy label design that uses the stairs rating in many EU countries might not be understood by Japanese consumers. For example, although the label with the stairs rating allows comparison, it is likely that the Japanese respondents had to make judgement that they are usually unaccustomed to making. On the contrary, since they did not have to make decisions based on such judgement for the label with the rating scale, they would not show any bias towards this label.

5. Conclusion

This study takes two label designs for energy savings into consideration: one with a stairs rating and one with a rating scale, and then uses survey findings to compare how these two types of labels are understood and related to energy use evaluation. The effects of the different types of labels on energy conservation decisions is also analyzed through a logistic regression analysis. This study assumes that energy labels, which are used to inform consumers on the energy efficiency level of houses, can motivate consumers to choose housing with lower levels of energy consumption.

The logistic regression analysis shows that when respondents are greatly concerned about the environment, they are easily motivated towards housing that offer higher levels of energy efficiency. The results also suggest that residents of colder regions are more likely to take energy conservation into consideration for housing. These results contrast with the results of those who are unemployed. In other words, there are differences in motivating individuals towards choosing housing with higher levels of energy saving depending on the respondent attributes. The results which exclude the control variables of the logistic regression analysis show that there is a problem with the label that has the stairs rating: the Japanese respondents have difficulties in grasping the information.

The anchoring effect is amplified by an increase in the level of the reference point, but not by higher standards, regardless of the label design. Furthermore, the anchoring effect shows evidence that the label with the rating scale makes it easier to choose an option that would incur higher energy savings as opposed to the label with the stairs rating. While the latter is expected to be easier to understand due to the information printed on the amount of energy consumed and the vertical shape, the results indicate otherwise.

The cross-tabulation shows that even though the label with the rating scale causes a slightly higher confusion rate compared to that of the label with the stairs rating, it is generally preferred. In addition, this result is supported by a higher stray rate of the label with the stairs rating. Thus, the label with the rating scale is preferred, not only based on the results of the logistic regression analysis, but also based on the respondents who show a preference for this label, which suggests that the label with the rating scale would encourage buyers to choose housing with higher energy savings. That is, the label with the rating scale is easier for the Japanese to understand. However, this label still needs to show text information on the amount of energy consumed on the rating scale and use gradation colors that allow a better understanding of the amount of energy used.

We find that there is the likelihood that the framing effect of the reference point words changes the respondent behavior. A reference point that shows a legal standard value will reduce the confusion rate. This means that consumers might possibly tend to choose housing with higher levels of energy saving because they understand the evaluated standards and accept their civic responsibilities. Moreover, the reference point has an important repercussion, that is, when the level of the reference point is increased, there is the tendency of the respondents to choose housing with a higher energy efficiency. Since the home insulation performance level has been low in Japan, it can be expected that increasing the home insulation level would lead to more energy efficiency.

The results of this comparative study on energy label designs are novel in the literature. They suggest that modifications to energy label designs are cultural, so that Japanese specific energy labels should be mandated soon to increase energy conservation.

A limitation of this study is that the subjects are all Japanese. The results of this study suggest that the label with the stairs rating might not be well understood by Japanese consumers. Since many countries in the EU have already accepted and adopted this type of label, the results of this study should be reexamined based on an investigation of those countries. Another limitation is the uncertainty of whether the findings of this study will actually encourage consumers to choose energy-efficient houses. These issues are left for future studies.

Acknowledgment

The authors would like to thank the anonymous reviewer for valuable comments. The responsibility of the remaining errors is with the authors.

This study is part of a research project supported by the Grants-in-Aid from The Telecommunications Advancement Foundation (Research Director: Mieko Fujisawa, Research Period: 2017-2018).

References

Andor, M., Gerster, A. and Sommer, S. (2017). Consumer Inattention, Heuristic Thinking and the Role of Energy Labels. USAEE Working Paper, No. 16-287, 1-29.

Behavioural Insights Team. (2011). Behaviour Change and Energy Use. Cabinet Office U.K., 1-35.

Bjerregaard, C. and Møller, N.F. (2019). The Impact of EU's Energy Labeling Policy: An Econometric Analysis of Increased Transparency in the Market for Cold Appliances in Denmark, *Energy Policy*, 128, 891-899.

Blach, J., Filippini, M. and Kumar, N. (2019). Boundedly Rational Consumers, Energy and Investment Literacy, and the Display of Information on Household Appliances, *Resource and Energy Economics*, 56, 39-58.

Brounen, D. and Kok, N. (2011). On the Economics of Energy Labels in the Housing Market, *Journal of Environmental Economics and Management*, 62, 166-179.

Bucchianeri, G.W. and Minson, J.A. (2013) A Homeowner's Dilemma: Anchoring in Residential Real Estate Transactions, *Journal of Economic Behavior & Organization*, 89, 79-92.

Bull, J. (2012). Loads of Green Washing—Can Behavioural Economics Increase Willingness-to-Pay for Efficient Washing Machines in the UK? *Energy Policy*, 50, 489-497.

Charness, G., Gneezy, U. and Kuhn, M.A. (2012). Experimental Methods: Between-Subject and Within-Subject Design, *Journal of Economic Behavior & Organization*, 81-1, 79-92.

Fuerst, F. and McAllister, P. (2011). The Impact of Energy Performance Certificates on the Rental and Capital Values of Commercial Property Assets, *Energy Policy*, 39, 6608-6614.

Fuerst, F., McAllister, P., Nanda, A. and Wyatt, P. (2015). Does Energy Efficiency Matter to Home-Buyer? An Investigation of EPC Rating and Transaction Prices in England, *Energy Economics*, 48, 145-156.

Gillingham, K., Rapson, D. and Wagner, G. (2016). The rebound Effect and Energy Efficiency Policy. *Review of Environmental Economics and Policy*, 10(1), 68-88.

Heinzle, S.L. and Wustenhagen, R. (2012). Dynamic Adjustment of Eco-Labeling Schemes and Consumer Choice – The Revision of the EU Energy Label as a Missed Opportunity? *Business Strategy and the Environment*, 21, 60-70.

Kahneman, D., Slovic, P. and Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press: New York, NY.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk, *Econometrica*, 47, 2, 263-292.

Mandel, N. and Johnson, E.J. (2002). When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices, *Journal of Consumer Research*, 29, 235-245.

Murphy, L. (2014). The Influence of the Energy Performance Certificate: The Dutch Case, *Energy Policy*, 67, 664-672.

Murray, A.G. and Mills, B.F. (2011). Read the Label! Energy Star Appliance Label Awareness and Uptake among U.S. Consumers, *Energy Economics*, 33, 1103-1110

Northcaft, G.B. and Neale, M. A. (1987). Experts, Amateurs, and Real Estate: An Anchoring-and- Adjustment Perspective on Property Pricing Decisions, *Organizational Behavior and Human Decision Processes*, 39, 84-97.

Ölander, F. and Thøgersen, J. (2014). Informing versus Nudging in Environmental Policy, *Journal of Consumer Policy*, 37, 3, 341-356.

Plevin I., Schnittjer, S.K. and Thee, S.L. (1988). Information Framing Effects in Social and Personal decisions, *Journal of Experimental Social Psychology*, 24, 6, 520-529.

Serenko, A. and Bontis, N. (2013). First In, Best Dressed: The Presence of Order-Effect Bias in Journal Ranking Surveys, *Journal of Informetrics*, 7, 1, 138-144.

Stadelmann, M. and Schubert, R. (2018). How do Different Designs of Energy Labels Influence Purchases of Household Appliances? A Field Study in Switzerland, *Ecological Economics*, 144, 112-123.

Thaler, R.H. and Sunstein, C.R. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness*. Yale University Press: New Haven.

Thonipara, A., Runst, P., Bizer, K. and Ochsner, C. (2019). Energy Efficiency of Residential Buildings in the European Union: An Exploratory Analysis of Cross-Country Consumption Patterns, *Energy Policy*, 129, 1156-1167.

Tversky, A. and Kahneman, D. (1974). The Framing of Decisions and the Psychology of Choice, *Science*, 211, 4481, 453-458.

Van de Velde, L. Verbeke, W., Popp, M. and Van Huylenbroeck, G. (2010). The Importance of Message Framing for Providing Information about Sustainability and Environmental Aspects of Energy, *Energy Policy*, 38, 5541-5549.

Waechter, S., Sutterlin, B., Borghff, J. and Segrist, M. (2016). Letters, Signs, and Colors: How the Display of Energy-Efficiency Information Influences Consumer Assessments of Products, *Energy Research & Social Science*, 15, 86-95.

Waechter, S., Sutterlin, B. and Segrist, M. (2015). The Misleading Effect of Energy Efficiency Information on Perceived Energy Friendliness of Electric Goods, *Journal of Cleaner Production*, 93, 193-202.

Walls, M., Geraden, T., Palme, K. and Bak, X. F. (2017). Is Energy Efficiency Capitalized into Home Prices? Evidence from Three U.S. Cities, *Journal of Environmental Economics and Management*, 82, 104-124.

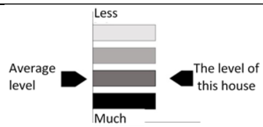
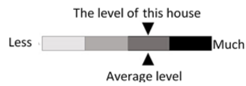
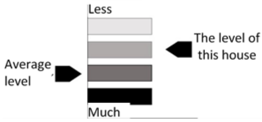
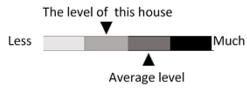
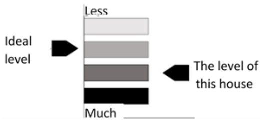
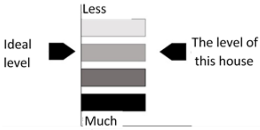
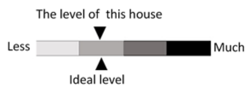
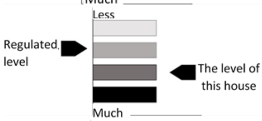
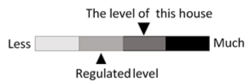
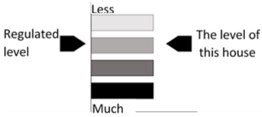
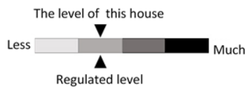
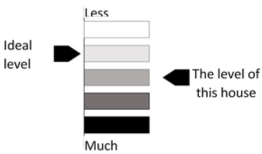
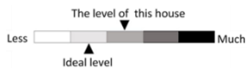
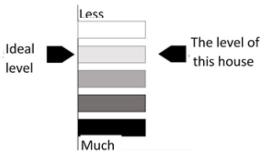
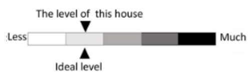
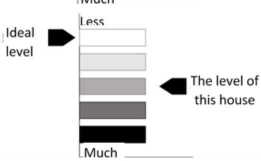
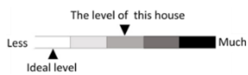
Wang, A. (2011). Priming, Framing, and Position on Corporate Social Responsibility, *Journal of Public Relations Research*, 19, 2, 123-145.

Ward, D.O., Clark, C.D., Jensen, K.L., Yen, S.T. and Russell, C.S. (2011). Factors Influencing Willingness-to-Pay for the ENERGY STAR® Label, *Energy Policy*, 39, 1450-1458.

Wiel, S. and McMahon, J.E. (2005). Energy-Efficiency Labels and Standards: A Guidebook for Appliances, Equipment, and Lighting. USDOE, 1-340.

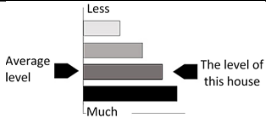
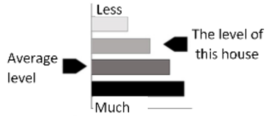
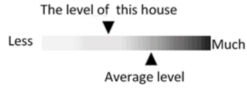
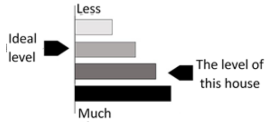
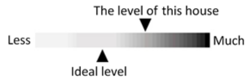
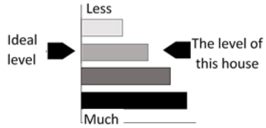
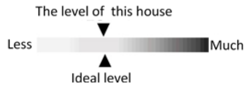
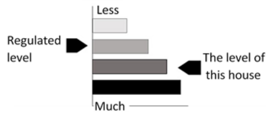
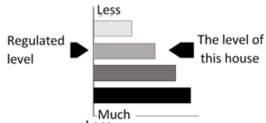
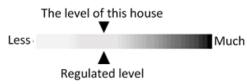
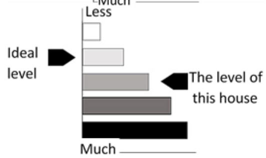
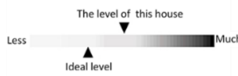
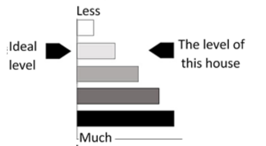
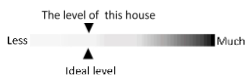
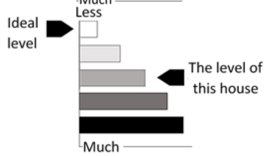
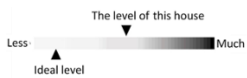
Appendices

Appendix A: Design Type Used for Each Survey Question

Question Number		Black & white, Same width	
Treatment A	Treatment B	Type 1	Type 2
Q1	Q10		
Q2	Q11		
Q3	Q12		
Q4	Q13		
Q5	Q14		
Q6	Q15		
Q7	Q16		
Q8	Q17		
Q9	Q18		

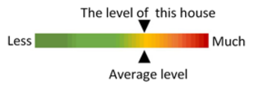
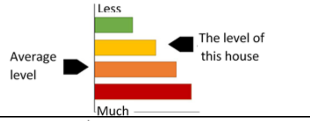
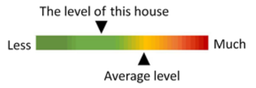
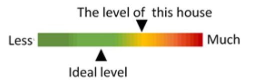
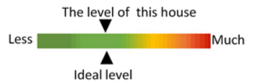
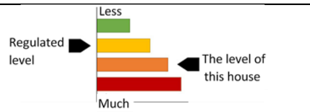
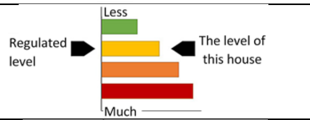
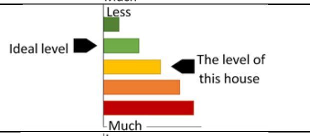
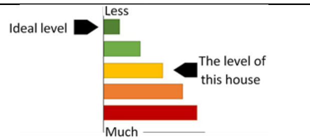
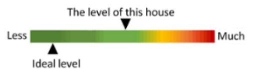
(Continued…)

(Appendix A Continued)

Question Number		Black & white, Same width	
Treatment A	Treatment B	Type 1	Type 2
Q1	Q10		
Q2	Q11		
Q3	Q12		
Q4	Q13		
Q5	Q14		
Q6	Q15		
Q7	Q16		
Q8	Q17		
Q9	Q18		

(Continued…)

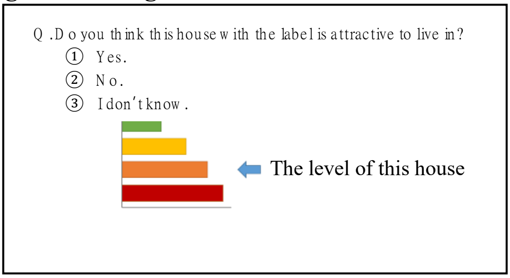
(Appendix A Continued)

Question Number		Black & white, Same width	
Treatment A	Treatment B	Type 1	Type 2
Q1	Q10		
Q2	Q11		
Q3	Q12		
Q4	Q13		
Q5	Q14		
Q6	Q15		
Q7	Q16		
Q8	Q17		
Q9	Q18		

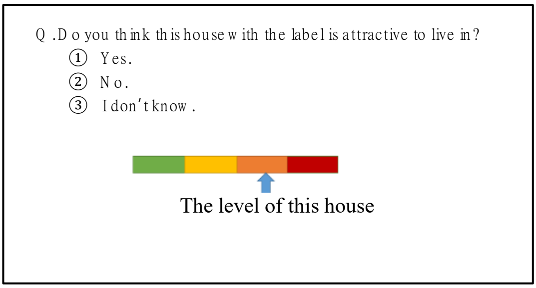
Appendix B Differences in Pilot and Online Surveys

1) Pilot Survey

Stairs Rating Form Design

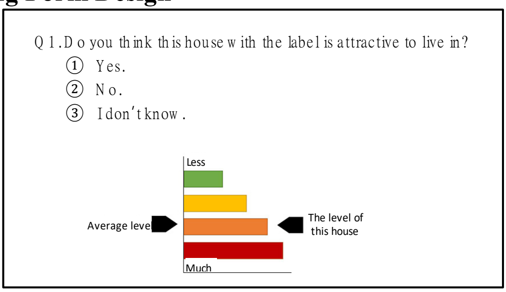


Rating Scale



2) Online Survey

Stairs Rating Form Design



Rating Scale