WINPEC Working Paper Series No.E2203
August 2022

Measuring the Response to Housing Energy Labels in Japan by Using an Eye-Tracking Experiment

Mieko Fujisawa, Kazuhisa Takemura, Yukihiko Funaki

Waseda INstitute of Political EConomy
Waseda University
Tokyo, Japan

Measuring the Response to Housing Energy Labels in Japan by Using an Eye-Tracking Experiment

Mieko Fujisawa[a*], Kazuhisa Takemura[b], Yukihiko Funaki[b]

Abstract

This study focuses on energy labels, which are set to be displayed mandatorily in Japanese real estate advertisements soon. In this study, we conducted eye-tracking experiments to identify effective design elements for energy labels. The novelty of this study lies in the fact that we not only collected data on reaction times and areas of interest (AOIs) using eye tracking, but also conducted a panel analysis controlling for individual effects by adding data from a questionnaire survey conducted after the experiment. Our findings verified that the display of energy labels in real estate advertisements is likely to lead to improved consumer understanding of energy conservation standards as learning effects. This suggests rehearsal effects that invited availability heuristics by appearing repeatedly. Moreover, the results of the panel analysis suggest that design of energy labels are important on reaction time and number of round trips between the AOIs. We compared the two label designs in the experiment, the information in the European Union energy label was difficult to read and judge intuitively, and can conclude the rating scale label was more suitable for advertising and readers in Japan. As energy labels help with increased consumer awareness regarding energy standards of dwellings and energy saving, an early start to labeling is recommended.

Keywords: energy label, eye-tracking, label design, response times, AOI, energy-saving policy

---

a Kanazawa University, Kakuma-machi, Kanazawa, Ishikawa, 920-1192, Japan.

E-mail: fujisawa@staff.kanazawa-u.ac.jp, *Corresponding author

b Waseda University

1. Introduction

To achieve the goals of sustainable residential areas and a reduction in carbon dioxide (CO_2) emissions within households, we must switch to high-performing housing insulation (energy-saving houses) from the prevalent low-insulation ones (Ministry of Land, Infrastructure, Transport and Tourism; MLIT, 2019). For consumers to choose and rent/buy energy-saving houses, energy labels are crucial in helping them make environment-friendly decisions.

Energy labels resemble environmental labels (eco-labeling) displayed to encourage the consumers who select green products used in energy-saving policies, such as those displayed on home appliances and real estate advertisements. In particular, energy labels on real estate advertisements are expected to encourage consumers to choose energy-saving houses. These energy labels will soon be mandatorily displayed in real estate advertisements to provide a crucial opportunity to the Japanese government to successfully reach its energy-efficiency goals. Therefore, it is very important to verify how effective such energy labels will be before they are mandated by the government.

This study aimed to clarify, by using an eye-tracking experiment, an effective design element of energy labels that makes it easier for consumers to read and quickly understand the energy consumption level of buildings. The advantage of the eye-tracking experiment is that it enables researchers to keep track of consumers' decision-making time and follow the line of sight of comparisons between labels. Further, it helps identify easy-to-read labels, as energy labels need to be recognized and understood readily by consumers for an optimum positive effect. The premise of this study is that easy-to-read labels lead to consumer understanding, thus encouraging them to choose energy-saving homes, and reduce their own, and the overall household energy consumption of Japan.

The novelty of our study is that we not only used the data collected from eye-tracking, similar to Brazil and Caulfied (2017), but also carried out an analysis with panel data to control individual effects, using both eye-tracking data and a questionnaire survey conducted after the experiment. Thus, our data made it possible to strictly measure the effect of energy label design, and we were able to extend the work done in previous studies, such as Stadelmann and Schubert (2018), which was using a cross-section analysis.

The remainder of this paper is organized as follows. Section 2 provides an overview of previous research and the position of this study. Section 3 describes the experiment setup and the analysis method, Section 4 describes the data, and Section 5 explains the results of the analyses and discusses effective label design. Section 6 concludes and presents policy implications regarding energy label systems.

2. Literature review

The approach to reducing energy consumption should not merely be through housing insulation repair; rather, a sociological approach should be employed as several studies have pointed out, including Sovacool (2014). One of the sociological approaches emphasizes the need for information that encourages consumers to change their behavior toward energy conservation. For example, as some studies have verified, the amount of energy consumption used as information provision changed consumers' actions on energy consumption (Heinzle & Wustenhagen, 2012), and utilizing energy management reports worked to stimulate their energy saving awareness (e.g., Allcott & Rogers, 2014; Fischer, 2008). Another approach is to conduct surveys on state-dependent and endogenous preferences. Bimonte, Bosco, and Stabile (2020) attempted WTP for eco-friendly products and verified that nudging is effective when it occurs during the decision process.

According to Ölander and Thøgersen (2014), considering previous research on the effect of information and the subsequent nudge to consumers, the use of both energy information and labels is deemed important; this study refers to extant research on energy information but focuses specifically on energy labels. In this section, we first discuss previous studies on energy labels and eye-tracking, and then explain the hypotheses of this study.

2.1 Study of energy labels

Countries consider energy labels to be a guide to energy conservation behavior, in terms of a nudge to consumers for an energy-saving effect (Behavioral Insights Team, 2011). Energy labels play a significant role, as is confirmed when consumers select a product after reading the information provided in the label, as an anchor or default reference point (e.g., Bucchianeri & Minson, 2013). Bjerregaard and Møller (2019) analyzed the before and after conditions of mandatory implementation of energy labels, and found that consumer behavior had indeed changed after the implementation of the mandate.

Several previous studies have examined consumers' pro-environmental behaviors such as choosing energy-saving products by looking at energy labels. Some have also verified the effectiveness of energy labels. Stadelmann and Schubert (2018) and Blach, Filippini, and Kumar (2019) have confirmed the positive effect of energy labels on appliances that facilitate simple calculations of expected energy usage.

Sammer and Wüstenhagen (2006) revealed the positive effects of energy labels on the consumers' willingness to pay (WTP) for home appliances (washing machines) in Switzerland by using a discrete-choice model. Shen and Saijo (2009) explored its effects on the use of air conditioners and refrigerators in Shanghai. Zhou and Bukenya (2016) conducted a discrete choice experiment in Germany and verified that efficient home appliances were selected when running costs were considered in combination with energy labels. Andor, Gerster, and Sommer (2017) examined the payment for energy-efficient products by using WTP and confirmed that consumers in Germany were willing to pay EUR 30 as an initial extra cost for energy-efficient refrigerators. Thus, the significant role of energy labels is confirmed when consumers select a product after checking out the label as a default reference point (Bucchianeri & Minson, 2013; Zhou & Bukenya, 2016). The same tendency is not only observed in research on home appliances, but also in studies on buildings. Stanley, Lyons, and Lyons (2016) verified that improved energy efficiency pushed up building prices, further, Lakić, Carroll, and Gubina (2021) also verified higher WTP for energy-efficient buildings.

However, studies on energy labels are sometimes not validated. In countries where energy labeling is mandatory, studies have shown that customer choices based on energy labels may not always be reasonable. Waechter, Sütterlin, and Siegrist (2015a) found that consumers chose larger refrigerators instead of selecting high-efficiency ones even after looking at the energy label, and in the end, their total energy consumption increased. In an experiment on energy labels, which involved adding annual operating cost information, Skourtos, Damigos, Tourkolias, and Kontogianni (2021) investigated the WTP for refrigerators, and no positive relationship was observed. Moreover, Thonipara, Runst, Ochsner, and Bizer (2019) discovered that the effects of energy labels were not constant across all countries where it was mandatory to have energy labels advertised, among other features, and that the changes were influenced more by the introduction of a carbon tax.

2.2 The eye-tracking experiment

Since energy labels encourage consumers' choice of energy-saving housing based on an appropriate evaluation, design research is important, because consumers must find the label's meaning easy to understand. Eye-tracking experiments with label design have been conducted on food information labels, and the superior effect of summarizing and transmitting information via labels has been verified by analyzing the time of attention fixation (Siegrist, Leins-Hess, & Keller, 2015).

Several studies have been conducted on the design of energy labels. In 2010, the European Union's (EU) energy label was revised to display not only a color scale but also an alphabetical rating scale, such as A to G. The evaluation of A was further fine-tuned to be more detailed, such as from A^{+} to A^{+++}. Heinzle and Wustenhagen (2012) analyzed the effects of the alphabetical scale display by comparing the old and new labels, however, they could not confirm an effect. Under rescaling of the current labels to return to an A to G scale, Boyano and Moons (2020) advocated the importance of energy labels to facilitate the adoption of dishwashers. Waechter, Sütterlin, Borghoff, and Siegrist (2016) focused on letters, signs, and colors of labels to examine which factors affected consumer choices, and suggested that the extended alphabetical scale raised the importance of energy efficiency, but consumers did not choose energy-efficient products after this change.

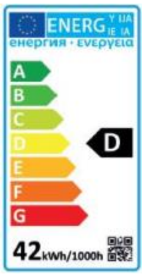

Fig. 1 The EU energy label

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

Fujisawa, Takemura, Funaki, Uto, and Takahashi (2020) carried out a web survey with a sample size of 1,078 regarding the design of energy labels (stairs rating-type and rating scale-type design), and confirmed the preference for Japanese rating scale-type labeling. The rating scale-type has been used as an energy label in Germany and other countries, and is called a tachometer or a continuum type due to its continuous shape design. However, the stairs rating-type is more common in the EU and is known as the EU energy label (Figure 1). On the other hand, the energy label of a continuum-type design with highly salient context information can affect real estate decision-making was as shown by Sussman, Conrad, Kormos, Park, and Cooper (2021), who carried out an experiment incorporating energy labels into real estate advertising.

Among design studies, Waechter, Sütterlin, and Siegrist (2015b) and Brazil and Caulfield (2017) used eye-tracking to verify the framing effect and design of the EU energy label (Table 1). Waechter, Sütterlin, and Siegrist (2015b) provided a systematic analysis of the EU energy label by using eye-tracking. This experiment was carried out on a sample of 117 people in Switzerland; it verified that the EU energy label can serve as a trigger for energy efficiency, and therefore, suggested a higher awareness of environmental considerations. At the same time, these results suggested that personal preferences for attributes were presumably much more important than energy-related information. Brazil and Caulfield (2017) focused on the elements that made up these labels and how effective they were in communicating the necessary information to the consumer. The results showed that the labels were designed in such a manner that there was a consistency of approaches across a number of sectors and greater potential for such labels to play a more prominent role in consumer choices, as proven by the experiment conducted on a sample of 43 students in Ireland.

Table 1: Summary of previous experiments

Study	Label	Subjects	Data	Analysis method
Waechter, Sütterlin, and Siegrist (2015b)	EU energy label	Consumers (N=117)	Fixation times / numbers Numbers of saccades	Two-way variance
Brazil and Caulfield (2017)	EU energy label	Students (N=43)	Heat maps Fixation times / numbers	Check of recollection tasks

2.3 Approach of this study

Our study conducted an experiment to compare two types of labels, based on the research by Waechter, Sütterlin, and Siegrist (2015b) and Brazil and Caulfield (2017), which only verified the EU energy label design but did not compare multiple energy labels. There are many graphic types of energy labels, such as the energy star in the United States, and these have been verified to have the same effect as the EU energy label (Murray & Mills, 2011; Ward, Clark, Jensen, Yen, & Russell,, 2011). Moreover, Sussman Conrad, Kormos, Park, and Cooper (2021) have proposed verification based on the real estate advertising experiment, such as detail factors analysis of energy labels design, as a future task.

Therefore, this study aimed to shed light on the superiority or inferiority of the design elements of multiple energy labels. This study focused on the EU options of stairs rating (Figure 2) and rating scale (Figure 3) designs because the Japanese government is presently trying to introduce an energy label with the graphic elements of the latter.

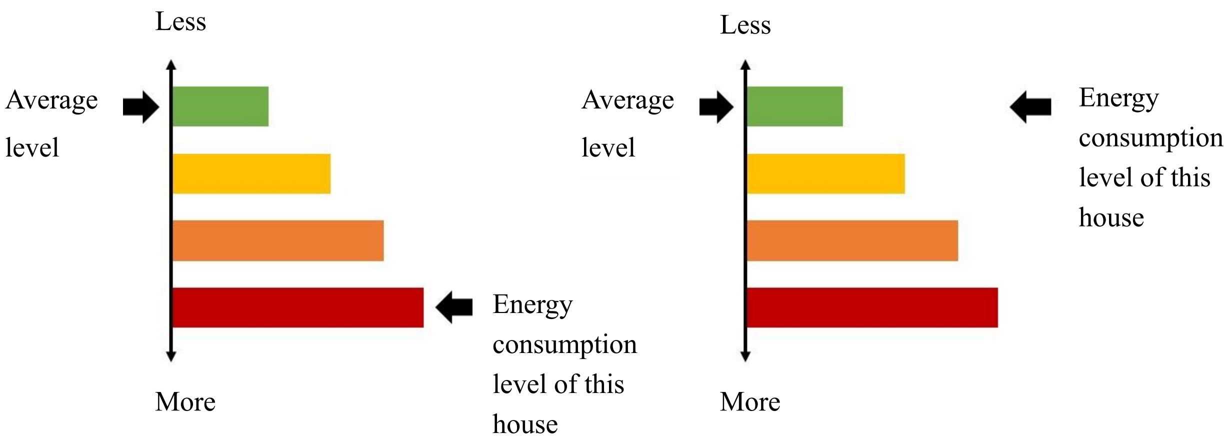

Fig. 2 The stairs rating-type design

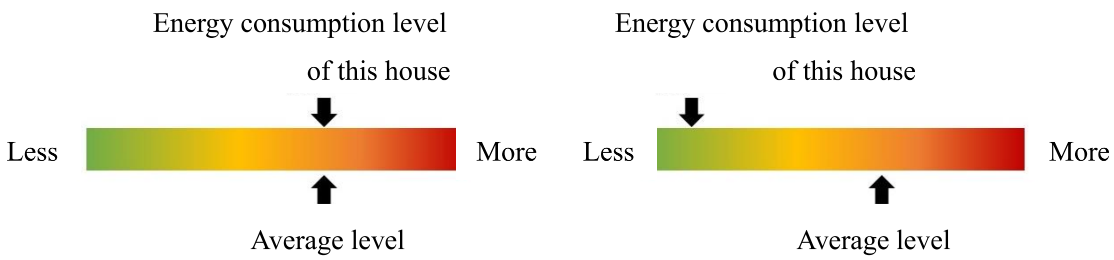

Fig. 3 The rating scale-type design

The research question for this study is following as: Do the design elements of energy labels make difference in consumers' understanding of energy-saving information? We assumed that easy-to-read labels would meet the following conditions: users would take a short time to make a decision (comprehensibility), and they would be able to judge the status of energy consumption at first glance (good visualization).

Thus, to clarify our research question we propose the following null hypotheses:

Hypothesis 1: There is no relationship between reaction time and the design of energy labels.

Hypothesis 2: The round trips for the comparisons of energy labels have no relationship with the design of the energy labels.

3. Methods

This section describes the experiment setup used in this study and the analysis methods that include ordinary least squares method (OLS) regression and panel analysis models.

3.1 Participants

The experiment was carried out in two stages: the pre-experiment[1] and the main experiment. The pre-

[1] The pre-experiment was conducted in September 2020, among the students of Hokuriku University, near

experiment checked for problems in operation and experiment time, while the main experiment was conducted over three days from October 19 to 21, 2020, at Kanazawa University. After the pre-experiment, we sent out an email to the students at Kanazawa University calling for experiment participants. The selected participants comprised of 35 university students (15 female and 20 male), with an average age of 19.57 years (SD: 1.20). The average time taken to conduct by this experiment was approximately 21 minutes. Although the number of participants was small in comparison to previous studies (Brazil & Caulfield, 2017), the experiment collected substantial data from each participant's slides, and sufficient samples were obtained for the empirical analysis using panel data as well.

All participants were first asked to read and sign a consent form that informed them that their gaze behavior would be recorded, but their data would be treated anonymously, and they could quit the study at any time without providing a reason. Next, the eye-tracking device structure was explained to them before the experiment could begin. The reward for each participant was 1,000 JPY ($11.04 on August 12, 2021), equivalent to twice the hourly wage for campus work. The participants were informed of the reward amount in the application email as well as in the consent form.

3.2 Eye-tracking device and software

We used an eye-tracking device and a software in this experiment. The former was used to collect data on eye movements and the latter for the aggregate data.

We used the GP3 Eye Tracker, a device made by Gazepoint, which has features such as a 60 Hz sampling rate, 1920 and 1080 resolutions, and a calibration of five points (four points are corners and one center point). We set this eye-tracking device in front of a PC monitor that the participants operated during the experiment. Using this device, we collected the reaction time and the number of round trips between two areas of interest (AOIs), that is, how many times the line of sight moved between AOIs, which were designed as shown in Figure 4. The reaction time was defined to measure the process time, from opening a slide to clicking the keyboard button, when participants made the decision to select the more energy-saving label out of the two labels in the slide.

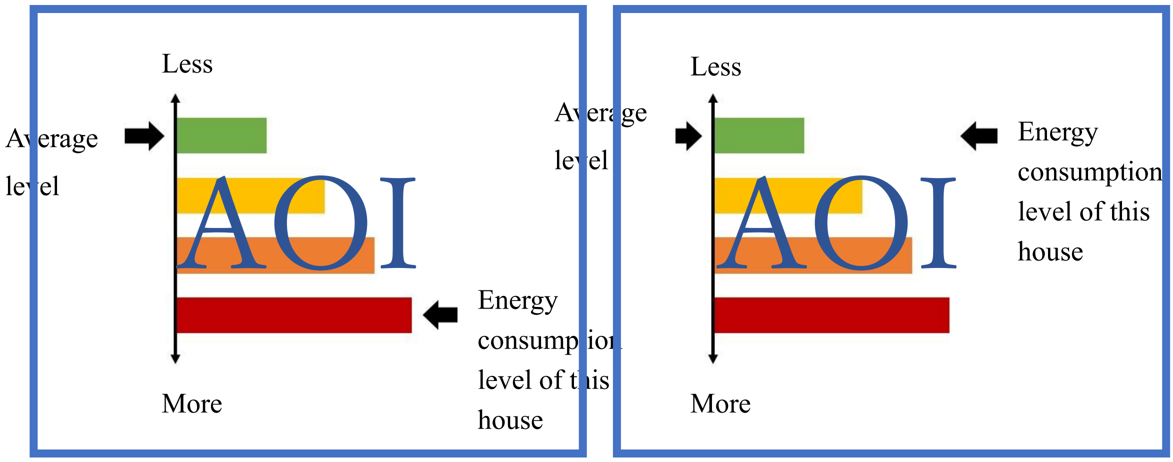

Fig. 4 AOI per slide

Kanazawa University in Ishikawa Prefecture.

We used the OGAMA version 5.0 software to collect and aggregate the eye-tracking data, after the reaction time was collected and the number of round trips on each slide counted. As the participants worked repeatedly at the slides (48 tasks), we collected 1,680 data points. Stata 16.0 statistical software was used for analyzing the data.

3.3 Experimental setup

This experiment adopted the between as well as the within method, because from the results of their experiment on labeling effects, Sörqvist, Haga, Holmgren, and Hansla (2015) suggested that comparisons should be made in the context of a between subjects-design.

To eliminate the sequential effects of using the between method for comparing the data between groups, the participants were divided into two groups: Group 1, in which 16 participants started with the stairs rating-type, and Group 2, in which 19 participants started with the rating scale-type.

Figure 5 shows the flow of this experiment. Each subject answered 48 questions and the total number of slides was 66, including explanatory slides. The participants were asked to choose the more energy-saving option in each slide. After the experiment, they answered a questionnaire survey on a PC monitor. This survey had 11 questions on various topics, including gender, age, easy-to-read label, favorite label, pro-environment behavior, and environmental knowledge; these were created based on the work done by Kollmuss and Agyeman (2002) and Walls, Gerarden, Palmer, and Bak (2017). Gender and Age were meaningful as control variables for demographic data, as confirmed by previous studies (Wang, Wang, & Guo, 2017). To distinguish between the easy-to-read and favorite labels of participants, the questionnaire survey asked both easy-to-read and favorite of them. Moreover, the pro-environment behavior and the environmental knowledge questions each had five sub-questions that were created based on the work of Ramos, Labandeira, and Löschel (2016).

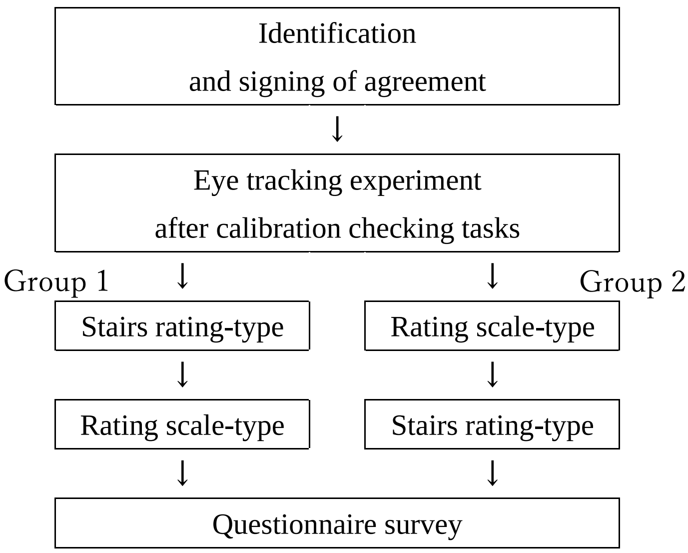

Fig. 5 The experiment flow

Owing to the use of the within method, two types of labels were tested on every participant in the experiment, thereby facilitating a holistic comparison between the stairs rating-types and rating scale-types. That is, as the participants of Group 1 responded to the slide questions of the rating scale-type after answering the stairs rating-type questions, they answered both label designs. In order to clarify the elements of the label designs, we devised the slides. Concretely, we designed this experiment such that participants could watch, in parallel, a slide containing two energy labels (stairs rating and rating scale), based on Schulte-Mecklenbeck, Sohn, de Bellis, Martin, and Hertwig (2013) (Figures 2 and 3). In addition, two patterns were created on both sides (left and right) of the reference label, based on Tourangeau, Couper, and Conrad (2004).

3.4 Analysis method

This study conducted its analysis in two dependent variables: Analysis 1 used the reaction time and Analysis 2 used the number of round trips between AIOs as the dependent variable. These were used for verifying Hypotheses 1 and 2, respectively. In this study, all the data were first pooled and then analyzed using OLS by following the model below:

$$Y_i = \alpha + \sum_{i=1}^{n} \beta_i X_i + \varepsilon \quad (1)$$

where Y represented the dependent variable, α the constant term, β the regression coefficient, X the independent variable, i the number of independent variables (i = 1,2, ..., n), and ε the error term.

Next, we carried out a panel analysis focusing on individual effects. The panel analysis control not only unobserved individual effects, including error terms ε, but also the learning effects. The model formula for the individual effect (A_i) was as follows:

$$Y_{it} = \alpha + \sum_{i=1}^{n} \beta_{it} X_{it} + A_i + \varepsilon \quad (2)$$

Here, Y, α, β, X, and ε had the same meaning as in formula (1). Furthermore, t represented the task numbers of this experiment (from 1 to 48). Similar to OLS, the actual estimating formula was as follows:

$$Y_{it} = \alpha + \sum_{i=1}^{n} \beta_{it} X_{it} + \varepsilon \quad (3)$$

In addition, we divided data into the stairs rating-type and the rating scale-type design data, because of comparing two designs. Thus, each design data was analyzed by each three models: OLS, fixed effects model, and random effect model.

4. Data

From the 1,680 samples (35 participants x 48 slides) generated from the experiment, we were able to collect 1,645 samples after excluding the ones with missing values.

4.1 Dependent variables

The two dependent variables of this study are reaction time and number of round trips between AOIs, Tables 2 and 3 describe the averages them on each Group, and show the results of each cross tabulation by chi-square test and t-test.

Table 2 shows the reaction times (second) of Groups 1 and 2 for each label. There was no difference between the energy labels, but we discovered that the responses to the labels viewed earlier was longer than those viewed later. There seemed to be some consumer learning effects, suggesting that if consumers get used to seeing energy labels, their reaction time would be cut short by about 30%.

Table 2: Reaction times

	Stairs rating-type	Rating scale-type
Group 1 (Starting with stairs rating)	3.856	2.637
Group 2 (Starting with rating scale)	2.711	4.041
Total average reaction time	3.235	3.399

Note: The t test confirmed the significance of the difference between Groups 1 and 2. Stairs rating-type: t (21.8) = 2.801, Rating scale-type: t (33) = -1.478.

From Table 3, it is clear that the number of round trips was fewer in the answers of the latter half than in the first half of the slides, which suggests that if consumers get used to seeing energy labels, their round trips would be reduce by around 12.3%.

Table 3: Number of round trips

	Stairs rating-type	Rating scale-type
Group 1 (Starting with stairs rating)	3.176	2.805
Group 2 (Starting with rating scale)	2.330	2.674
Total average reaction time	2.727	2.734

Note: The t test confirmed significance of the difference between Groups 1 and 2. Stairs rating-type: t (21.4) = 3.493, Rating scale-type: t (31) = 0.532.

4.2 Independent variables

The independent variables were based on three factors: experiment setup, reaction, and attribute. Table 4 shows the data summary of the independent variables (see Appendix for the correlation matrix of data).

The variable of experiment setup explained each screen situation on factors of base, type, and gap. The base was the benchmark picture that displayed the same reference point and energy-saving standard level. The variable of *Base left dummy* was the case that displayed the benchmark picture on the left. There were two types of design on the experiment screen (stairs rating-type and rating scale-type); the *Stairs rating-type dummy* means the case that displayed the stairs rating-type label. These two variables were dummy variables and were coded 1 if applicable. *Gap* meant the difference inside a label, between the arrow's standard level as the reference point and the arrow indicating the actual energy-saving level of an applicable housing.

The reaction factor was configured with three variables and two interaction variables. *Correct dummy* was coded 1, when the participants answered with the correct option from the two graphics of labels in a slide. *Easy-to-read label dummy* was coded 1, when the participants answered with their the easiest-to-read option for the stairs rating-type label. Similarly, *Favorite label dummy* was coded 1 when the participants answered with their favorite option for the stairs rating-type label. In addition, we used two cross-terms as interaction to accurately measure the effects of label design: *Easy-to-read SR dummy* was the intersection of stairs rating-type dummy and easy-to-read label dummy, and *Favorite label SR dummy* was the intersection of stairs rating-type dummy and favorite label dummy.

The attribute factor had five variables. If the participant is a man, *Gender* was coded 1. *Age* is a real number. *Pro-environmental points* referred to the sum of participants who answered yes to pro-environment behavior in question. Similarly, *Environmental knowledge* referred to the sum of subjects who answered well to the environmental questions, and *Experience dummy* referred to the participants experience with housing contracts in the past.

Table 4: Summary of data

Variable	Mean	Std. Dev.	Min	Max
Reaction time	3342.675	2202.319	444	17,758
Numbers of round trips	2.4696	1.7374	0	11
Experiment setup factor				
Base left dummy	0.4996	0.5002	0	1
Stairs rating-type dummy	0.4990	0.5002	0	1
Gap	1.6620	0.7432	1	3
Group 1 dummy	0.4571	0.4983	0	1
Reaction factor				
Correct dummy	0.9392	0.2390	0	1
Easy-to-read label dummy	0.4261	0.4947	0	1
Favorite label dummy	0.4833	0.4999	0	1
(Interaction)				
Easy-to-read SR dummy	0.6571	0.4748	0	1
Favorite SR dummy	0.6571	0.4748	0	1
Attribute factor				
Gender	0.5714	0.4950	0	1
Age	19.5714	1.2024	18	22
Pro-environmental points	4.2571	1.6453	2	8
Environmental knowledge	2.4	0.5453	1	3
Experience dummy	0.7429	0.4372	0	1

5. Results

After reporting the results of Analysis 1 and 2, and we discuss the label design based on these results. The result of each Analysis explains from three models: first model was analyzed using all data, next model used only stairs rating-type data, and last model used only rating scale-type data. We confirm the results of all data case at first, and we interpret difference with each two designs data.

5.1 Estimation results of Analysis 1

The results of Analysis 1 estimates are shown in Table 5. In addition, all the variance inflation factors (VIF) to confirm multi-collinearity were less than 10 (Mean VIF was 1.50). An F-test was performed to verify whether individual effects could be significantly detected after the analysis (F (14, 1630) = 11.25, P <0.001). All null hypotheses without individual effects were then rejected (F test result's Prob > F = 0.0000). Next, to determine whether individual effects were correlated with explanatory variables, a Hausman test was applied to compare the fixed effects and random effects models (Houseman test result's Prob > chi2 = 0.0547). As a result, the random effects model was adopted. Hence, the following focuses primarily on the results of the random effects model.

The short impact variables for the participants' reactions time were *Base left dummy*, *Gap*, *Correct dummy, Easy-to-read SR dummy*, and *Favorite SR dummy*. In particular, the *Easy-to-read SR dummy* was statistically significant by 1% and had a large impact. The results of *Base left dummy* signified the benchmark painting label that participants found easy to read. Similarly, the participants found the label with the bigger *Gap* between the benchmark and comparison easy to read. Moreover, it suggests that the correct answer was intuitively selected from the result of the *Correct dummy*. The results were interesting, in that the participants were able to answer their *easy-to-read label* in a short reaction time. However, the variables of *Stairs rating-type dummy* elicited a longer reaction time among participants. As an exception, if the participants like the Stairs rating-type, it could be deemed to have the effect of shortening the reaction times from the result of *Easy-to-read SR dummy*.

Table 5: Estimated results of Analysis 1

Dependent variable	Time (All data)			Time (Stairs rating-type data)			Time (Rating scale-type data)		
	OLS	Fixed effects	**Random effects**	OLS	**Fixed effects**	Random effects	OLS	Fixed effects	**Random effects**
Base left dummy	-181.1372 *	-186.2951 *	-185.8857 *	-67.9341	-102.1157	-97.2232	-262.2624 **	-259.8621 **	-260.0333 **
	104.1655	95.3107	95.34461	135.0783	120.8878	121.3262	147.9299	130.3716	130.1945
Stairs rating-type dummy	203.4816 *	254.4548 **	250.2623 **						
	119.3924	110.0258	109.9959						
Gap	-126.7255 *	-119.5191 *	-120.0981 *	-116.0004	-106.8026	-108.1132	-130.5502	-125.3261	-125.7314
	70.1358	64.1903	64.2117	90.9110	81.34687	81.6440	99.3444	87.6401	87.5085
First half SR dummy	-172.9124	(omitted)	-161.4942	978.2821 ***	(omitted)	987.5143 ***	-1580.4180 ***	(omitted)	-1570.4950
	117.3543		376.2241	149.1074		358.4398	168.1604		475.6590
Correct dummy	-1357.7410 ***	-1344.08 ***	-1345.696 ***	-2287.8180 ***	-1737.1130 ***	-1816.7320 ***	-675.7691 **	-751.7424 **	-739.7492 **
	230.3698	232.01	230.1723	480.0121	440.6035	440.6043	268.4376	318.3324	306.7601
Easy-to-read label dummy	-507.0705 ***	(omitted)	372.4018	305.0750	(omitted)	291.3071	354.8691	(omitted)	351.0530
	150.5421		497.2472	197.4467		473.3257	219.0025		626.5273
Favorite label dummy	-75.3454	(omitted)	-1106.309 **	-1184.0770 ***	(omitted)	-1158.5660 **	-1156.3790 ***	(omitted)	-1159.4700 **
	134.7392		521.4194	206.9596		497.4537	227.8025		657.6002
Easy-to-read label ×SR	361.1389 **	-587.1112 ***	-580.1124 ***						
	156.9633	148.1696	147.3027						
Favorite label ×SR	-1114.2930 ***	-120.273	-117.0183						
	160.4850	128.6217	128.1998						
Gender	-366.2308 ***	(omitted)	-365.2932	-391.2664 **	(omitted)	-388.1993	-465.6266 **	(omitted)	-457.5490
	137.3301		445.2158	176.7760		424.6833	196.1979		562.1028
Age	54.9672	(omitted)	55.13293	46.3279	(omitted)	48.9564	49.7489	(omitted)	48.0704
	51.7305		168.3643	66.8952		160.6679	73.4020		212.3030
Pro-environmental points	80.2538 **	(omitted)	81.19384	-29.5102	(omitted)	-29.8186	194.3899	(omitted)	191.8488
	38.1359		123.3881	48.8269		117.6178	54.4628		155.7745
Environmental knowledge	280.7436 **	(omitted)	280.7696	641.1781 ***	(omitted)	638.9948 **	32.4300 ***	(omitted)	29.0943
	123.0621		397.2493	158.1257		378.9145	175.3006		501.6181
Experience dummy	366.4445 ***	(omitted)	367.9247	651.8814 ***	(omitted)	647.9459	133.7901	(omitted)	129.6963 ***
	130.8043		424.053	168.2555		404.4451	186.8698		535.4047
Constant term	3489.7620 ***	5078.1000 ***	3470.914	3266.1420 **	5190.3400 ***	2750.6580	3903.3150 ***	4438.4670 ***	4001.9130
	1080.4430	249.5045	3384.41	1452.0990	456.6645	3257.5570	1519.2560	344.2768	4271.3780
Sample size	1,645	1,645	1,645	821	821	821	824	824	824
Adj R-squared	0.0803	0.0523	0.0523	0.1704	0.0231	0.0231	0.1347	0.0137	0.0137

Note: 1) Superscripts ***, **, * denote significance at the level of 1%, 5%, and 10%, respectively.

2) All data: F test: Prob > F = 0.0000, Houseman test: Prob>chi2 = 0.0547, Breusch and Pagan test: Prob > chibar2 = 0.0000

3) Stairs rating-type data: F test: Prob > F = 0.0000, Houseman test: Prob>chi2 = 0.0000, Breusch and Pagan test: Prob > chibar2 = 0.0000

4) Rating scale-type data: F test: Prob > F = 0.0000, Houseman test: Prob>chi2 = 0.9978, Breusch and Pagan test: Prob > chibar2 = 0.0000

Different models and variables have been adopted in the analysis of each data, respectively. For the data of stairs rating-type, the fixed effects model was adopted, suggesting the magnitude of the individual effects due to individual differences. Conversely, the random effects model was adopted for the rating scale-type, *Base left dummy* was statistically significant on this data only; this seems to be a point that must be kept in mind designing the rating scale-type. The variables had statistically significant through all models were *Correct dummy* and *Favorite dummy*.

5.2 Estimation results of Analysis 2

The results of all the model estimates are shown in Table 6. In addition, all of the VIF to confirm multi-collinearity were less than 10 (Mean VIF was 1.51). An F-test was performed to verify whether individual effects could be significantly detected after the analysis ($F (14, 1533) = 10.36$, $P <0.001$). All null hypotheses without individual effects were then rejected (F test result's Prob > F = 0.0000). Next, to determine whether individual effects were correlated with explanatory variables, a Hausman test was applied to compare the fixed effects and random effects models (Houseman test result's Prob > chi2 = 0.9997). As a result, the random effects model was adopted. Hence, the following results focus primarily on the random effects model results.

The variables that decreased the numbers of round trips were *Base left dummy*, *Gap*, *Correct dummy*, and *Easy-to-read SR dummy*. In particular, *Correct dummy* was statistically significant at the 1% level and had a high impact. On the other hand, the variables of the *Stairs rating-type dummy* and *Group 1 dummy* had an increased positive impact on the number of round trips among participants. This suggests that the stairs rating-type label is not participant-friendly, and the stairs design did not appeal to them.

The random effects model was adopted on both analysis which divided into separately each design data. As the statistically significant variables were different, only in stairs rating-type was confirmed statistical significance of *Correct dummy*, suggesting that it had intuitive design features. However, stairs rating-type showed plus influence for number of round trips judging from the coefficients of *Stairs rating-type dummy* and *First half SR dummy*. In generally, the stairs rating-type design tends to increase the number of round trips, and it seems difficult for anyone other than accepting its design to understand. Therefore, we judged the stairs rating-type design is unacceptable to all consumers.

Table 6: Estimated results of Analysis 2

Dependent variable	Number of round trips (All data)			Number of round trips (Stairs rating-type data)			Number of round trips(Rating scale-type data)		
	OLS	Fixed effects	**Random effects**	OLS	Fixed effects	**Random effects**	OLS	Fixed effects	**Random effects**
Base left dummy	-0.2396***	-0.2349***	-0.2351***	-0.184294	-0.202801*	-0.201192*	-0.281343**	-0.265964**	-0.267633**
	0.0848	0.0771	0.0770	0.1211571	0.1088823	0.1088125	0.1184307	0.1049574	0.1048476
Stairs rating-type dummy	0.3407***	0.3403***	0.3405***						
	0.0956	0.0874	0.0872						
Gap	-0.2197***	-0.2197***	-0.2198***	-0.159379**	-0.15454**	-0.155393**	-0.279892***	-0.275436***	-0.276781***
	0.0572	0.0521	0.0520	0.0815503	0.0732598	0.0732164	0.0798686	0.0708268	0.070745
First half SR dummy	0.6932***	(omitted)	0.7296*	0.9267033***	(omitted)	0.9923548***	0.2968918**	(omitted)	0.2922358**
	0.0988		0.3848	0.1382315		0.3763299	0.1396396		0.3958312
Correct dummy	-0.8777***	-0.7960***	-0.8036***	-1.16947***	-0.903534**	-0.933796**	-0.839694***	-0.499302**	-0.56212
	0.1839	0.1838	0.1825	0.4182016	0.3846888	0.3834591	0.2121041	0.2520198	0.2429752
Easy-to-read label dummy	0.1944	(omitted)	0.1462	0.0307631	(omitted)	-0.026616	0.1802277	(omitted)	0.1694905
	0.1264		0.4881	0.1752619		0.4769941	0.1730647		0.5004012
Favorite label dummy	-0.3879***	(omitted)	-0.3506	-0.593257***	(omitted)	-0.572182	-0.250563	(omitted)	-0.210324
	0.1299		0.5171	0.1848998		0.5058843	0.1811362		0.5301229
Easy-to-read label ×SR	-0.5135***	-0.5114***	-0.5114***						
	0.1199	0.1173	0.1166						
Favorite label ×SR	0.0154	-0.0331	-0.0306						
	0.1079	0.1024	0.1020						
Gender	-0.1506	(omitted)	-0.1624	-0.195576	(omitted)	-0.26441	-0.178948	(omitted)	-0.177487
	0.1148		0.4481	0.1639332		0.4396044	0.1604569		0.4599611
Age	-0.0098	(omitted)	0.0126	-0.010181	(omitted)	0.0049979	-0.018882	(omitted)	0.0105089
	0.0415		0.1649	0.0591572		0.1613722	0.0579072		0.169094
Pro-environmental points	-0.0075	(omitted)	-0.0072	0.0437771	(omitted)	0.0357941	-0.073279*	(omitted)	-0.062885
	0.0303		0.1200	0.0426667		0.1173105	0.042648		0.1232026
Environmental knowledge	0.1200	(omitted)	0.1568	0.3337574**	(omitted)	0.4052532	-0.058733	(omitted)	-0.02033
	0.0999		0.3930	0.1421378		0.3850017	0.1391979		0.4033531
Experience dummy	0.0087	(omitted)	-0.0577	0.1949457	(omitted)	0.1631838	-0.183365	(omitted)	-0.233306
	0.1128		0.4414	0.1589347		0.4317972	0.1595203		0.4539557
Constant term	3.6428***	3.7603***	3.0866	3.111495**	3.744324***	2.469311	4.770461***	3.473582***	3.794459
	0.8615	0.1977	3.2905	1.276585	0.3996736	3.242727	1.190073	0.2719033	3.381699
Sample size	1,548	1,548	1,548	772	772	772	776	776	776
Adj R-squared	0.0781	0.0545	0.0545	0.1192	0.0188	0.0137	0.0445	0.0321	0.032

Note: 1) Superscripts ***, **, * denote significance at the level of 1%, 5%, and 10%, respectively.

2) All data: F test: Prob > F = 0.0000, Houseman test: Prob>chi2 = 0.9997, Breusch and Pagan test: Prob > chibar2 = 0.0000

3) Stairs rating-type data: F test: Prob > F = 0.0030, Houseman test: Prob>chi2 = 0.7721 Breusch and Pagan test: Prob > chibar2 = 0.0000

4) Rating scale-type data: F test: Prob > F = 0.0000, Houseman test: Prob>chi2 = 0.8104, Breusch and Pagan test: Prob > chibar2 = 0.0000

5.3 Discussion

Cross tabulation results revealed that the learning effects shortened the reaction time for selection and reduced the number of round trips for comparison. This result is consistent with Berenger and Møller (2019), who verified that energy labeling was mandatory and obligatory, and that the positive effect was seen only after normalization. Moreover, this suggested rehearsal effects that invited availability heuristics by appearing repeatedly. Given the possibility that the display of energy labels on real estate advertisements as an energy-saving policy may lead to consumers' understanding of energy-saving standards; the results in Zhou and Bukenya (2016) too showed similar trends.

All analysis were adopted fixed effects or random effects model, and it had become to increase the possibility of strict identification without individual effects of the variables that affected the reaction time or the number of round trips, thus, making it expanded on the study by Waechter, Sütterlin, and Siegrist (2015b). In addition, results of this study due to examine them labeling effect through comparing two designs of labels from two rigorous method (between and with-in methods) could update the experiment by Sörqvist, Haga, Holmgren, and Hansla (2015).

Hypothesis 1 was rejected by the results of Analysis 1 and an alternative hypothesis, which states that the readability design of label promoted faster responses, was suggested. Similarly, an alternative hypothesis 2, which states that a label with good visualization can lead to intuitive judgment and is a factor that reduces the number of round trips, was also verified by the results of Analysis 2. Simultaneously, the results of the analysis revealed that good visualization and design elements are necessary for consumers to evaluate energy labels properly. Specifically, we found that it was better to place the information or energy standard on the left side of the energy labels, in particularly rating scale-type. This result is the same as that of Tourangeau, Couper, and Conrad (2004), which also verified that users found the left side of a slide easier to read. Moreover, we found that a large gap between the average energy consumption level and the energy consumption level of a particular house within a label enabled a faster decision. Allcott and Rogers (2014) show that a benchmark sign in a label for similar households in the surrounding area promotes behavioral change in consumers toward energy-saving, especially when the difference between households is emphasized in the visual elements of the label. It seems worthwhile to visualize the difference between standard and real housing insulation performance on the energy label, as an operational aspect. As for the effect of visualization, since the coefficient of *correct dummy* variable shows reduced reaction times and numbers of round trips, there is a clear possibility of intuitive judgment of information through the energy labels. This information transmission effect of energy labels had been also verified from the experimental results of Sussman, Conrad, Kormos, Park, and Cooper (2021).

Based on the experimental results, we can conclude that the rating scale-type label is more suitable for advertising. As the results of dummy variables of Stairs rating-type and first half SR suggest, the stairs rating-type label caused difficulty in understanding or confusion because the reaction times and the number of round trips between AOIs increased. This finding was in line with the findings of Fujisawa, Takemura, Funaki, Uto, and Takahashi (2020); furthermore, our study was able to confirm this trend in greater detail, by verifying that the trend of reaction time was similarly for the number of round trips between AOIs. From the above, it is considered that a continuous label such as the rating scale-type is optimal in Japan, rather than the stairs rating-type, which is common in EU. We speculate that continuum-type design may be suitable for Japanese, for the reason, one of the result due to social science approach pointed out by Sovacool (2014), it is that there is some cultural background, but, another detail survey might be needed to verify this result.

In addition, we found that reading labels is not affected by gender, age, and environmental interests or knowledge. This differs from the results of previous studies (Walls, Gerarden, Palmer, and Bak, 2017), even though this study focused on the design of energy labels and did not confirm the relationship between labels and consumer action. Therefore, the study still needs to verify whether consumers actually choose energy-saving houses based on the energy labels advertised.

6. Conclusion

This study focused on the design elements and visibility of energy labels that will soon be mandatorily to be displayed in real estate advertisements in Japan. We conducted an eye-tracking experiment to verify the best suitable label design. From results of panel analysis, the alternative hypotheses were adopted and confirmed that comprehensibility and good visualization are important for energy labels' design.

Displaying the energy labels of houses when consumers decide to buy or rent a house, will promote the selection of energy-saving houses. Considering the vital role of the energy labels here, we find two policy implications.

First, the energy labeling of houses for sale or rent needs to be mandated to provide an opportunity for consumers to consider energy saving. Thus, consumers will become accustomed to the energy labels and be able to make faster energy-saving decisions when purchasing or renting real estate. While waiting for legal obligations, governments can also adopt a policy to promptly promote labeling, even if optional.

Second, the rating scale-type label seemed desirable enough to be adopted and displayed in real estate advertising. Adopting this energy label will lead to greater intuitive judgment, promoting consumers' choice of energy-saving houses by examining the energy label. For this purpose, it is important to place the reference information on the left side of the label when displaying labels. If the energy-saving level of the house is poor, the label design should emphasize how different it is from the energy-saving standard.

These proposals need to be tested and verified in the future with further experiments conducted on the energy labels displayed in actual advertisements. The study has two limitations regarding the experiment slides and participants. First, this experiment did not reflect real-life labels because the slides were created for assessing design elements. Second, the participants were students, who do not generally buy houses, limiting the generalizability for considering housing decisions. All these issues need to be addressed in future research.

Acknowledgment

We would like to thank Nobuyuki Uto for his help with our experiment and suggestions on the experiment design.

This study was supported by the Japan Society for the Promotion of Science (JSPS) through a Grand-in-Aid for Scientific Research (KAKENHI) 19K01604 (Research Director: Mieko Fujisawa, Research period: 2019–2022).

Ethics approval

This experiment followed the "Kanazawa University Researcher Behavior Standards" (enacted January 22, 2008); Certification Number: 02-29.

References

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, *104*(10), 3003–3037. https://doi.org/10.1257/aer.104.10.3003

Andor, M., Gerster, A., & Sommer, S. (2017). *Consumer inattention, heuristic thinking and the role of energy labels [USAEE working paper]*, http://www.rwi-essen.de/ media/content/ pages/publikationen/ruhr-economic-papers/rep_17_671.pdf, *671* (pp. 1–34)

Behavioral Insights Team. (2011). Behaviour change and energy use. Cabinet Office U.K. https://www.bi.team/publications/behaviour-change-and-energy-use-behavioural-insights-team-paper/ (pp. 1–35).

Bimonte, S., Bosco, L., & Stabile, A. (2020). Nudging pro-environmental behavior: Evidence from a web experiment on priming and WTP. *Journal of Environmental Planning and Management*, *63*(4), 651–668. https://doi.org/10.1080/09640568.2019.1603364

Bjerregaard, C., & 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. https://doi.org/10.1016/j.enpol.2019.01.057

'Blasch, J., Filippini, M., & 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. DOI: 10.1016/j.reseneeco.2017.06.001

Boyano, A., & Moons, H. (2020). Analyzing the impacts of the revision of ecodesign and Energy Label Regulations for household dishwashers using a stock model. *Energy Efficiency*, *13*(6), 1147–1162. https://doi.org/10.1007/s12053-020-09874-4

Brazil, W., & Caulfield, B. (2017). What makes an effective energy efficiency label? Assessing the performance of energy labels through eye-tracking experiments in Ireland. *Energy Research and Social Science*, *29*, 46–52. https://doi.org/10.1016/j.erss.2017.05.014

Bucchianeri, G. W., & Minson, J. A. (2013). A homeowner's dilemma: Anchoring in residential real estate transactions. *Journal of Economic Behavior and Organization*, *89*, 76–92. https://doi.org/10.1016/j.jebo.2013.01.010

Fischer, C. (2008). Feedback on household electricity consumption: A tool for saving energy? *Energy Efficiency*, *1*(1), 79–104. https://doi.org/10.1007/s12053-008-9009-7

Fujisawa, M., Takemura, K., Funaki, Y., Uto, N., & Takahashi, R. (2020). An experimental study of the effect of energy label design on the correct evaluation of buildings energy performance and promotion of energy-saving. *International Real Estate Review*, *23*(1), 107–136. http://dx.doi.org/10.53383/100296

Heinzle, S. L., & Wüstenhagen, 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*(1), 60–70. https://doi.org/10.1002/bse.722

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally, and what are the barriers to pro-environmental behavior? *Environmental Education Research*, *8*(3), 239–260. https://doi.org/10.1080/13504620220145401

Lakić, E., Carroll, J., & Gubina, A. F. (2021). The role of monetary information in energy performance certificates: A Slovenian experiment. *Energy Efficiency*, *14*(7), 1–17. https://doi.org/10.1007/s12053-021-09995-4

Ministry of Land, Infrastructure, Transport and Tourism, & MLIT. (2019). Chapter 8. Creating and preserving a beautiful and healthy environment. https://www1.mlit.go.jp/ common/001325170. pdf. In *White paper on land, infrastructure, transport and tourism in JAPAN 2019*, *1*: Promoting global warming countermeasures.

Murray, A. G., & Mills, B. F. (2011). Read the label! Energy Star appliance label awareness and uptake among US consumers. *Energy Economics*, *33*(6), 1103–1110. https://doi.org/10.1016/j.eneco.2011.04.013

Ölander, F., & Thøgersen, J. (2014). Informing versus nudging in environmental policy. *Journal of Consumer Policy*, *37*(3), 341–356. https://doi.org/10.1007/s10603-014-9256-2

Ramos, A., Labandeira, X., & Löschel, A. (2016). Pro-environmental households and energy efficiency in Spain. *Environmental and Resource Economics*, *63*(2), 367–393. https://doi.org/10.1007/s10640-015-9899-8

Sammer, K., & Wüstenhagen, R. (2006). The influence of eco-labelling on consumer behaviour- Results of a discrete choice analysis for washing machines. *Business Strategy and the Environment*, *15*(3), 185–199. https://doi.org/10.1002/bse.522

Schulte-Mecklenbeck, M., Sohn, M., de Bellis, E., Martin, N., & Hertwig, R. (2013). A lack of appetite for information and computation. Simple heuristics in food choice. *Appetite*, *71*, 242–251. https://doi.org/10.1016/j.appet.2013.08.008

Shen, J., & Saijo, T. (2009). Does an energy efficiency label alter consumers' purchasing decisions? A latent class approach based on a stated choice experiment in Shanghai. *Journal of*

Environmental Management, *90*(11), 3561–3573. https://doi.org/10.1016/j.jenvman.2009.06.010

Siegrist, M., Leins-Hess, R., & Keller, C. (2015). Which front-of-pack nutrition label is the most efficient one? The results of an eye-tracker study. *Food Quality and Preference*, *39*, 183–190. https://doi.org/10.1016/j.foodqual.2014.07.010

Skourtos, M., Damigos, D., Tourkolias, C., & Kontogianni, A. (2021). Efficient energy labelling: The impact of information content and style on product choice. *Energy Efficiency*, *14*(6), 1–19. https://doi.org/10.1007/s12053-021-09950-3

Sörqvist, P., Haga, A., Holmgren, M., & Hansla, A. (2015). An eco-label effect in the built environment: Performance and comfort effects of labeling a light source environmentally friendly. *Journal of Environmental Psychology*, *42*, 123–127. https://doi.org/10.1016/j.jenvp.2015.03.004

Sovacool, B. K. (2014). What are we doing here? Analyzing fifteen years of energy scholarship and proposing a social science research agenda. *Energy Research and Social Science*, *1*, 1–29. http://dx.doi.org/10.1016/j.erss.2014.02.003

Stadelmann, M., & 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. https://doi.org/10.1016/j.ecolecon.2017.07.031

Stanley, S., Lyons, R. C., & Lyons, S. (2016). The price effect of building energy ratings in the Dublin residential market. *Energy Efficiency*, *9*(4), 875–885. https://doi.org/10.1007/s12053-015-9396-5

Sussman, R., Conrad, S., Kormos, C., Park, C., & Cooper, E. (2021). Context and meaningfulness in energy efficiency labeling: Real estate listings. *Journal of Environmental Psychology*, *78*. https://doi.org/10.1016/j.jenvp.2021.101681

Thonipara, A., Runst, P., Ochsner, C., & Bizer, K. (2019). Energy efficiency of residential buildings in the European Union–An exploratory analysis of cross-country consumption patterns. *Energy Policy*, *129*, 1156–1167. https://doi.org/10.1016/j.enpol.2019.03.003

Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, *68*(3), 368–393. https://doi.org/10.1093/poq/nfh035

Waechter, S., Sütterlin, B., Borghoff, J., & Siegrist, M. (2016). Letters, signs, and colors: How the display of energy-efficiency information influences consumer assessments of products. *Energy Research and Social Science*, *15*, 86–95. https://doi.org/10.1016/j.erss.2016.03.022

Waechter, S., Sütterlin, B., & Siegrist, M. (2015a). The misleading effect of energy efficiency information on perceived energy friendliness of electric goods. *Journal of Cleaner Production*, *93*, 193–202. https://doi.org/10.1016/j.jclepro.2015.01.011

Waechter, S., Sütterlin, B., & Siegrist, M. (2015b). Desired and undesired effects of energy labels—An eye-tracking study. *PLOS ONE*, *10*(7), e0134132. https://doi.org/10.1371/journal.pone.0134132

Walls, M., Gerarden, T., Palmer, K., & Bak, X. F. (2017). Is energy efficiency capitalized into home prices? Evidence from three US cities. *Journal of Environmental Economics and Management*, *82*, 104–124. https://doi.org/10.1016/j.jeem.2016.11.006

Wang, Z., Wang, X., & Guo, D. (2017). Policy implications of the purchasing intentions towards energy-efficient appliances among China's urban residents: Do subsidies work? *Energy Policy*, *102*, 430–439. https://doi.org/10.1016/j.enpol.2016.12.049

Ward, D. O., Clark, C. D., Jensen, K. L., Yen, S. T., & Russell, C. S. (2011). Factors influencing willingness-to-pay for the ENERGY STAR® label. *Energy Policy*, *39*(3), 1450–1458. http://dx.doi.org/10.1016/j.enpol.2010.12.017

Zhou, H., & Bukenya, J. O. (2016). Information inefficiency and willingness-to-pay for energy-efficient technology: A stated preference approach for China Energy Label. *Energy Policy*, *91*, 12–21. https://doi.org/10.1016/j.enpol.2015.12.040

Appendix

	Y1	Y2	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14
Reaction time (Y1)	1															
Number of round trips (Y2)	0.548	1														
Base left dummy (X1)	-0.051	-0.067	1													
Stairs rating-type dummy (X2)	0.011	0.015	-0.004	1												
Gap (X3)	-0.028	-0.090	-0.004	-0.010	1											
First half SR dummy (X4)	-0.001	0.163	-0.008	0.030	0.005	1										
Correct dummy (X5)	-0.187	-0.088	0.016	0.186	-0.038	0.150	1									
Easy-to-read label dummy (X6)	-0.051	-0.059	0.009	-0.024	-0.006	-0.093	-0.090	1								
Favorite label dummy (X7)	-0.137	-0.124	0.011	-0.043	-0.011	-0.209	-0.110	0.593	1							
Easy-to-read SR dummy (X8)	-0.147	-0.098	0.003	0.430	0.000	0.121	0.189	0.202	0.135	1						
Favorite SR dummy (X9)	-0.117	-0.012	0.001	0.307	0.000	0.025	0.172	-0.005	0.056	0.577	1					
Gender (X10)	0.027	0.035	-0.008	-0.002	0.002	-0.071	0.029	-0.402	-0.408	-0.092	-0.076	1				
Age (X11)	0.046	0.026	-0.014	0.009	0.007	0.180	0.055	0.042	-0.223	0.049	-0.039	0.186	1			
Pro-environmental points (X12)	-0.004	0.018	0.001	0.022	0.004	0.159	-0.124	-0.008	0.268	0.056	0.013	-0.121	-0.351	1		
Environmental knowledge (X13)	0.097	0.008	-0.001	0.019	-0.009	-0.222	-0.091	-0.329	-0.219	-0.108	0.002	0.531	0.128	0.030	1	
Experience dummy (X14)	0.013	0.036	-0.008	0.001	0.007	0.115	-0.087	0.044	0.183	0.041	-0.058	0.187	0.114	0.182	0.000	1