Nudging Energy Conservation in University Housing: A Field Experiment in Taiwan

Ya-Wen Tseng* Jie-Yu Yang† Eric S. Lin‡ Yating Chuang§

February 4, 2023

Abstract

This paper describes a non-price-based conservation experiment performed among 6,723 college dorm residents. We analyze the effectiveness of a behavioral intervention embodied in an email sent to residents featuring one of two types of social comparison based energy consumption feedback: quintile social comparison message and mean comparison message. A quintile social comparison message ranks residents' relative energy consumption in quintiles, while the mean comparison message compares residents' energy consumption with the average. Our results show that the relative effectiveness of these messages depends on the residents' baseline consumption pattern. The highest 20% of users are more likely to reduce their energy consumption with a quintile comparison message, while the lowest 20% are more responsive to a mean comparison message. We also observe that messages are more effective with undergraduate residents than they are with graduate students. Our results suggest that feedback design should be context-specific.

1 Introduction

Social comparisons can affect our choices in various domains, such as financial decisions (Beshears et al., 2012; Breza, 2012), smoking and drinking (Nakajima, 2007; Eisenberg et al., 2014; Yakovlev, 2018), voting (Gerber and Rogers, 2009), energy use (Allcott, 2011; Allcott and Rogers, 2014), water use (Bhanot, 2017), home improvement decisions (Bollinger et al., 2012; Bollinger et al., 2020), and so on. It is popular for policymakers to use social comparisons as "nudge" style interventions to affect individuals' choices, an approach that has implications for private benefits and social welfare. Social comparison is useful for providing a reference point for individuals, yet it is often unclear what reference group people should be compared with. Do people care most about

*Department of Economics, National Tsing Hua University
†Department of Economics, National Tsing Hua University
‡Department of Economics, National Tsing Hua University
§Corresponding author, Institute of Economics, Academia Sinica, yating@econ.sinica.edu.tw

comparing themselves to the average? Or, do they care more about whether they are in the higher percentile vs. lower percentiles? Determining how to make social comparison nudges more effective has important policy implications.

We study a so-called "home energy report (HER)" intervention, a social-comparison-based nudge and widely studied program. The typical HER provides individuals with information regarding their energy usage compared to their neighbors along with some conservation tips. A company called Opower began this type of intervention, as it would be hired by utilities to improve operaional efficiency. When Opower began this intervention in the US in 2009, it was cost effective to reduce energy consumption by 2% (Allcott, 2011). Since then, HER has been widely implemented at 85 utilities in the United States, and similar HERs have been carried out in nine countries (Allcott and Kessler, 2019). Utilizing an HER intervention in university dormitories in Taiwan, we conducted a natural field experiment (Levitt and List, 2009), one in which the subjects were unaware that they were in a study. We sought to answer the following questions. First, are the comparative feedback messages used effective in this context? Second, what type of reference group comparison is more effective, and for whom is it more effective? Third, does sending the message a second time produce differing effects? To answer these questions, we randomize the messages sent using two social comparison references: quintile comparison and mean comparison. And we note the ways in which people respond to these messages differently. We also randomize the number of messages sent. Randomizing the messages helps fine-tune the social comparison with different reference points to better understand their effectiveness.Randomizing the number of messages helps us to better understand the channels that drive effective nudges. It could be that the effective nudge provides the essential energy usage information for people who did not know it. In this case, sending messages once or twice may result in the same conservation behavior. Alternatively, it could be that people have limited attention. This notion resonates with Akerlof (1991)'s and Bordalo et al. (2013)'s papers about how salience affects consumers' choices. If this mechanism dominates, sending messages twice would help to remind people to save energy more than sending them once.

We implemented our experiment in the dormitories at National Tsing Hua University (NTHU) in Taiwan in the fall semester of 2019. Rather than recruiting voluntary respondents, we have access to all the residents' emails at NTHU, allowing us to minimize selection bias and enhance our study's sample. With the school's permission, we sent designed messages to all students living

in the university dorms through the central administrative system. In the emails, we provide comparative feedback on students' electricity consumption. Our messages employed one of two types of framing. One used a quintile comparison where we inform residents of the quintile of their electricity consumption (*Qt group).* The other uses a mean comparison where residents are informed whether their consumption is lower or higher than the average (*Avg group*). Based on the framing design (*Qt group or Avg group*) and how many times the messages were sent (once or twice), we divided our residents into five randomization groups: the control group, the OnceQt group, the TwiceQt group, the OnceAvg group, and the TwiceAvg group.

Our results show that the conservation effect is greater when a heavier user (users with higher baseline electricity consumption) receives a quintile comparison. Conversely, a lighter user (users with low baseline electricity consumption) saves more when receiving a mean comparison message. The reason may be that a quintile comparison, relative to a mean comparison, provides more precise information, allowing heavier users to realize their outsized electricity consumption relative to their peers. For lighter users, a mean comparison may minimize a boomerang effect — a common concern in the literature that lighter utility users may increase energy consumption after receiving social comparison information. The comparative message was shown to affects the energy use of college students but not of graduate students. This result may reflect the varying economic status of different types of students and their relative time spent in the dormitory. Overall, messages sent more than once were more effective than those sent only once. This result suggests that salience and limited attention theory may be indicative factors explaining individual's energy saving behaviors.

Our research makes several contributions. First, our study demonstrates the replicability and scalability of HER conservation experiments (Allcott, 2015; Al-Ubaydli et al., 2017a,b; Andor et al., 2020; Dehejia and Linos, 2015; Camerer et al., 2018; Peters et al., 2018; Vivalt, 2015). Because of various common problems with experiment design, such as publication bias, non-adoption, non-adherence of treatments, site selection bias, and representativeness of the experimental population and situation, scholars have called for field experiments to be designed to be scalable and incorporate policy implications. Replicability is of great importance, as an open collaboration at the journal Science has found in replicating 100 experimental and correlational studies in three psychology journals and concluding that "*Replication can increase certainty when findings are reproduced and promote innovation when they are not*" (Open Science, 2015). This conclusion applies to the HER literature, which commonly features inconclusive results due to small and unrepresentative samples

and incompatible experiment settings. Beginning with the foundamental Allcott (2011) and Allcott and Rogers (2014) studies, the literature has generally shown HER interventions to be useful in many contexts, yet there are also exceptions. Myers and Souza's (2020)'s HER experiment at UIUC dorms does not show a reduction in energy consumption. Andor et al. (2020) conducted a large-scale HER field experiment in Germany and found that the treatment effect of HER is smaller than the magnitude found in US-based studies. They conclude that such a program may have less cost-effectiveness potential in Europe than elsewhere. Some researchers find that HER interventions can be enhanced when combined with other incentives. Researchers have, for instance, displayed university dorms' electricity usage comparisons in public, and this approach has led to 17% $\sim$ 20% increase in energy savings (Delmas and Lessem, 2014; Bruelisauer et al., 2018). Other researchers use monetary rewards to amplify social comparisons by hosting a competition to encourage participants to save energy and have found the added incentive effective (Brewer et al., 2011; Chen et al., 2021). This paper adds to the literature with evidence from a non-western country, an important contribution. Our experiment is close to a natural experiment, as we have studied the whole university population, and participants were not aware that they were being part of a study. Because of our collaboration with the schools' administration, we are less likely to suffer disadvantages related to selection bias.[1]

This study also contributes to the literature by better understanding the mechanism that drives the effectiveness of a nudge-style message, and, because we conducted a natural-style experiment, we contend that our results are exceptionally robust. Employing social comparison messages with different reference groups, our study points to ways of making these types of messages more effective. Social comparison messages have shown to effectively reduce people's energy consumption, yet it is unclear at what reference level this comparison works best. Scholars have documented the importance of designing feedback with optimal reference groups in both laboratory and natural experiments (Carrell et al., 2013; Kahneman and Tversky, 1979; Kuhnen and Tymula, 2012). The recent study by Roels and Su (2022) provides a theoretical framework highlighting design mechanisms for choosing appropriate reference groups. Our field experiment compliments their purely theoretical analysis by providing empirical guidance for designing the optimal social comparison. Optimal framing of social comparison messages, is, however, a qualitative metric, depending largely

[1]Most of the studies mentioned above are through recruitment except Allcott (2011) and Allcott and Rogers (2014)'s original Opower HER studies and Andor et al. (2020)'s study where they collaborate with a German's service provider for utilities.

upon the context of each situation and the group under examination. Research has shown, for instance, that providing conservation information related to health is especially effective in achieving energy savings among families with children – more than twice as effective as it is for the total sample (Asensio and Delmas, 2015). Costa and Kahn (2013) also find intriguing heterogeneity in the results among people with different political identities: energy conservation nudges are more effective among political liberals than among political conservatives.

We tested the relative effectiveness of the number of messages sent, seeking to add to the literature on salience and attention and determine, what role these factors play in behavioral shifts associated with HER intervention. The literature is not conclusive on this subject. No matter how often the messages are provided (daily, weekly, or monthly), some studies show significant effects, while others show insignificant effects (Allcott, 2011; Asensio and Delmas, 2015; Asensio and Delmas, 2016; Fischer, 2008; Meub et al., 2019; Myers and Souza, 2020). Yet there is a growing consensus that people may need sufficient time to adjust their electricity consumption habits to achieve a significant power-savings (Fischer, 2008; Allcott and Rogers, 2014; Anderson et al., 2017). One-shot information is not enough to encourage conservation behaviors (Fischer, 2008). This (in)effectiveness discussed in the literature may be related to salience — people's decision-making stems from the more salient information in the choice set. Researchers have found, for instance, that limited attention can explain people's saving behaviors (Karlan et al., 2014). Our experimental design yields results that describe whether salience or other mechanisms play roles in effecive nudging from the HER intervention. We hypothesize that salience has a role in decision-making if households act upon the second message more than the first. If the first-time message is more effective than the repeated message, then, it would suggest habituation as found in Allcott and Roger (2014).[2] If the two messages are similarly effective, then it points to information as catalyst — residents needed this information to decide whether or not to engage in conservation. To the best of our knowledge, we are the first to test whether the number of times messages are sent matters. These results could help policymakers to encourage significant behavioral change more cost-effectively.

Our study makes several contributions to the literature: First, we provide additional empirical evidence for behavioral based energy conservation beyond the US context. Second, we mitigate

[2]However, Ito et al. (2018) use a more refined experiment to test habituation/dishabituation and find that using moral suasion to encourage conservation will have quicker habituation than an economic incentivized program.

the common recruitment selection bias through a natural field experiment with universal access to all residents' emails. Last, our design helps better understand the mechanisms behind social comparison's effectiveness by examining different reference groups framing and presenting different levels of salience. These designs have policy-making implications.

2 Experiment Design

We conducted a randomized control trial (RCT) in the student dormitories at National Tsing Hua University. The University's delivered energy report provides data on individual's an opportunity to reveal the relative electricity usage compared with neighbors. Our experiment examines the subsequent behavioral change following a specific nudge. This section expands on the experiment design.

Unlike most electricity experiments in dormitories, we cooperated with the administrative unit to include all residents in our experiment without the usual recruitment process. In total, there are 6,895 residents distributed in twenty buildings and 2,709 rooms in the final sample. [3] There are four room types: single, twin, triple, and quadruple rooms.[4]

Considering the university's academic schedule, the climate, and the electricity consumption pattern in Taiwan, we conducted the experiment from September 12, 2019, to October 31, 2019. This timing follows the university's semester schedule – the fall semester in Taiwan starts in the middle of September and ends in early January. Electricity consumption in Taiwan is largely seasonal. Air conditioning is usually the primary source of energy consumption, while heating is rarely used in Taiwan. There is almost no heating in the dormitories. So our experiment is conducted during the season when air conditioning is heavily used. We decided that mid-September to October is the optimal period for our study because November in Taiwan is relatively cool.[5]

During the study period, electricity usage feedback messages were delivered by email bi-weekly, with two versions employed: a quintile comparison and a mean comparison. The former describes individual's energy usage in quintiles: top 20% as high users, bottom 20% as low users, and the

[3] Here is how the buildings are classified according to student status and gender: Thirteen of the dormitories are for undergraduates, three are for graduate dormitories, and the remaining are for mixed-use. Regarding gender, there are six female-only buildings, nine male-only buildings, and five mixed-gender buildings. The overall building complex goes from the second to tenth floors, while most buildings have three or four stories.

[4] There are 248 single rooms, 1,522 twin rooms, 153 triple rooms, and 786 quadruple rooms.

[5] Bruelisauer et al.'s (2018) experiment demonstrates that providing feedback of high-energy-use appliances could promote graduate students' energy conservation more effectively than providing general energy usage feedback. This finding supports taking high-energy-consuming appliances into account.

intermediate users as 21-40%, 41-60%, and 61-80% (see Figure 1). The latter depicts electricity consumption relative to the average, so residents are informed whether their consumption is higher or lower than the mean (see Figure 2). These messages convey to residents their energy consumption relative only to their student dormitory and building of residence.

We include facial expression emojis in both versions to help recipients understand the information more quickly and clearly since past studies indicate that such an expression could indicate social appropriateness for certain behaviors and reduce the boomerang effect, especially among lighter energy users (Schultz et al., 2007; Bonan et al., 2020). A survey link was sent with the electricity report to residents to get feedback on their intentions after receiving the message and the length of time they spend in their room.

The number of energy reports sent is a manipulated intervention in our experiment. With the same message content, one treatment group received the feedback twice while the other received the message once. These varying treatment would determine whether the number of messages would affect conservation behaviors. For this reason, the length of the experiment is divided into three periods: Baseline, Period I, and Period II. The treated information is sent in Periods I and II. The first three weeks comprise the Baseline, and the duration of the follow-up periods is two weeks.

Based on message framing and the number of messages sent, residents in the twenty buildings are randomly assigned into five groups by floor in each building: the control group, the once-quintile group (*OnceQt*), the twice-quintile group(*TwiceQt*), the once-average group(*OnceAvg*), and the twice-average group(*TwiceAvg*) (see Figure 3). This design is intended to balance the distribution of the five groups across floors and buildings to reduce the possible bias caused by differences between student halls and floors. Table 1 shows the five group's characteristics comparing dormitory types, students' backgrounds, and energy consumption.

3 Data and Methods

3.1 Data

To conduct this experiment among all dormitories on campus, we had to obtain a detailed resident list of all dorms and the electricity record for each room. We were granted through special permission detailed and reliable administrative-level residential data collected by the Division of Student Housing (a unit responsible for managing and operating university dormitories). The

administrative-level data comprises the student directory of all dorms, such as residents' department, gender, degree, grade, and email account. Regarding electricity record, the Office of General Affairs only has available the electricity usage per building. To observe residents' behavior, we collected room-level data by hand, observing electricity usage via the Watt-hour meter installed outside each room. The Watt-hour meter records air-conditioner usage charged by air-conditioner cards (mainly for 220V).[6] We hired eight part-time students to collect electricity records for each room every two to three weeks from September to October 2019.

Besides the student directory and usage data, we also conducted online surveys through a link in the emails sent to each resident in the treatment group. This survey helps us gather information on students' attitudes and behaviors toward electricity usage. We asked them, for example, if they would reduce their electricity consumption after receiving the social comparison message. The detailed questionnaire is in Appendix A. We designed this questionnaire with only five questions to increase the response rate. Answering the survey also entered respondents in a lottery, in which twenty-five entrants would win an NT$100 gift voucher.

Detailed information on the important demographic and electricity usage variables is shown in Table 1. We generate a layered individual-level dataset by combining three sources of data: the student directory of the dormitories, electricity records, and online surveys. The data set contains various important variables, such as the electricity usage per person per week, room characteristics (e.g., the located building/floor of the rooms and the number of residents per room), residents' characteristics (e.g., gender, degree, grade, and whether undergrad or not), and attitudes and behaviors related to energy conservation (see Appendix A). We exclude the information from rooms with incorrect electricity usage records, rooms with no residents, and those without students' email information. After this data-cleaning process, the number of rooms in the sample decreased from 2,709 to 2,668, and the total sample size reduces from 6,895 residents to 6,723 residents.

3.2 Descriptive Statistics

Table 1 shows the sample mean of the whole individual-level data, the sample mean by treatment group, and the results of the multivariate means test. Panel A of Table 1 presents the electricity usage of each experimental period. We can see that the mean usage in period II is the smallest, with

[6] It is worth noting that what is charged via air-conditioner cards is different by building. For eight buildings, residents pay for the electricity usage of the air-conditioner, fans, and lighting. In comparison, residents in the other twelve buildings only need to pay for the air-conditioner.

only 2.227 kWh per person per week and the smallest standard deviation. This trend is consistent with the temperature pattern during the experiment. As shown in Figure 4, the temperature in period II is more stable and lower than in the Baseline and Period I. Panel B of Table 1 presents the room characteristics. The average number of residents in a room is around 3. Panel C of Table 1 describes residences' characteristics. In the sample, 58.2% are male. Most residents are undergraduate students, accounting for 70.3% of the sample. Most of the residents are first-year and second-year students. The last column of Table 1 shows the p-values of tests examining whether the means across our control and four treatment groups are the same. As shown in the last column, all p-values are larger than 0.1, failing to reject the null hypothesis. This result validates our randomization process. Regarding the treatment effect, the last column of Panel A shows that the average usage across five groups does not have significant differences after the first and second treatments, suggesting that the nudging policy may not have had significant effects. This preliminary evidence, however, is insufficient to conclude the policy effect. We use a panel DID regression model to evaluate the treatment effect.

Table 2 presents summary statistics of our two surveys. The first survey was sent to all four treatment groups (N=5,283), and the response rate was 28.13%. The second survey was sent with the second treatment information; thus, only two treatment groups (N=2,590) received the second survey. The response rate of the second survey was 23.17%. We also asked respondents to fill out their usage ranking to see if their answers aligned with our records. This question helps us find out whether residents have read the message carefully - they are actually "being treated." In the first survey, 92.2% correctly provided their usage ranking. In the second survey, 87.5% answered correctly. Panel B in Table 2 compares the respondents' characteristics with the full sample. In our survey, females were more likely to respond to the questionnaires. The distribution of respondents' degrees was similar to the full sample, while the proportion of master students was higher in the survey. Panel C shows the results of the survey questions. First, we want to know students' willingness to change their electricity behavior after receiving feedback. According to Table 2, around 40% of them would like to reduce their electricity usage. Second, over 86% of those surveyed reported that they care about climate change or global warming in daily life. Finally, the average time they spent in the dormitory (after deducting sleep time) was around 5 hours per day. The mean time they stay in the dormitory per week is 5.685 days.[7]

[7]For the selection options of the average time spent in the dormitory, 1.5 represents less than 3 hours, 5 is 4 to 6

3.3 Estimation Methods

Our estimation model employs a Panel Difference-in-Differences Method (DID) to evaluate the effect of the two proposed nudging policies. As in Section 2, we randomly assigned dormitory rooms into five groups. One is the control group, while the other four are treatment groups that vary with feedback design and the number of messages sent. Because the experiment lasts for two months and we collect the electricity usage every two to three weeks (four times in total), we organize the data into a panel data set to help control for time-invariant individual effects, such as gender, grades, and other educational backgrounds. The general form of DID in the panel data framework is as follows:

$$Y_{it} = \alpha_i + \lambda_t + \gamma Post_t + \beta Treat_i + \delta(Treat_i \times Post_t) + \varepsilon_{it}, \tag{1}$$

where Y_{it} is the outcome variable of resident i at time t. Specifically, in this study, Y_{it} is the electricity usage per person per week for resident i between the time point t and $t-1$. We divide the usage by the number of residents in their rooms - there are between one and four residents in each room. Because the frequency of collecting electricity records changes from every two to three weeks during the experiment, we transform the usage unit into a one-week unit. $Treat_i$ equals 1 if the entity i is assigned to the treatment group and 0 otherwise. $Post_t$ equals 1, representing the period after the policy, and 0 before the policy. The coefficient on the interaction term $Treat_i \times Post_t$ captures the causal treatment effect. The model also includes fixed effects, α_i, to control for resident fixed effects, and λ_t to control for time fixed effects. Finally, ε_{it} is the error term. In this main specification, δ is the parameter of interest, which measures the effects of electricity consumption induced by our message. We expect it to be negative if the policy is effective.

To determine the effectiveness of the feedback, we designed a two by two experiment - four treatment groups varied by message type (quintile vs. average comparison) and number of times message was sent (once vs. twice) (see Figure 3). In the Baseline (between $t-1$ and t), we have not yet messaged any residents. The first treatment – mean or quintile social comparison message - was sent out at time t. Except for the control group, the treatment groups were provided with two types of information: one is the quintile ranking of an individual's baseline electricity usage (Qt group), and the other is the comparison with the average usage of the baseline electricity consumption (Avg group). Among all the treated groups, we randomly select some of them to receive the second

hours, 8 is 7 to 9 hours, and 11 is more than 9 hours.

treatment (the message time treatment), where they receive the second-time information at time $t+1$.

There are four treatment groups in total: once-quintile group ($OnceQt$), twice-quintile group ($TwiceQt$), once-average group ($OnceAvg$), and twice-average group ($TwiceAvg$). The $OnceQt$ group and $TwiceQt$ group received the quintile ranking of their baseline electricity usage, and the $OnceAvg$ group and $TwiceAvg$ group got the messages in comparison with the baseline average electricity usage. At time $t+1$, only the $TwiceQt$ and $TwiceAvg$ groups received the second-time information of their Period I electricity usage.

We combine some groups and reorganize the time period to estimate the treatment effects among different messaging formats. In doing so, we cut the time period of our experiment into two overlapping sections. As shown in Figure 3, the first section contains the baseline and Period I (from $t-1$, t, to $t+1$), and the second section includes Period I and Period II (from t, $t+1$, to $t+2$).

During the first section, only the first treatment is implemented. Therefore, we combine the $OnceQt$ and $TwiceQt$ groups into one group, named $QtTreat$, as they receive the same quintile message treatment. Similarly, we put the $OnceAvg$ and $TwiceAvg$ groups together and name them $AvgTreat$ since they are exposed to the mean comparison message treatment. Our estimation is as follows:

$$\begin{aligned} Y_{it} = \alpha_i + \lambda_t + \gamma Post1_t + \beta_1 QtTreat_i + \delta_1(QtTreat_i \times Post1_t) \\ + \beta_2 AvgTreat_i + \delta_2(AvgTreat_i \times Post1_t) + \varepsilon_{it}, \end{aligned} \tag{2}$$

where $QtTreat_i$ equals 1 if the entity i is assigned to the $QtTreat$ group (including $OnceQt$ and $TwiceQt$, the quintile information group), and 0 otherwise. $AvgTreat_i$ equals 1 if the entity i is assigned to the $AvgTreat$ group (including $OnceAvg$ and $TwiceAvg$, the average information group), and 0 otherwise. The control group is the group without any feedback. $Post1_t$ equals 1, representing Period I, and 0 is the baseline. The coefficient on the interaction term $QtTreat_i \times Post1_t$ captures the treatment effect for the quintile message treatment. The coefficient on the interaction term $AvgTreat_i \times Post1_t$ captures the treatment effect for the mean comparison message treatment.

In the second section, we evaluate the difference between the first-time and the second-time treatment effect. We drop our original control group and treat the $OnceQt$ group as the control

group to compare with the $TwiceQt$ group. Since at time t, both the $OnceQt$ and $TwiceQt$ groups received the same quintile comparison information. While at time $t+1$, only the $TwiceQt$ group was exposed to the second mean comparison information. Following the same logic, we treat the $OnceAvg$ group as the control group of the $TwiceAvg$ group. By comparing the $TwiceQt$ with the $OnceQt$ group and the $TwiceAvg$ with the $OnceAvg$ group, we examine whether the number of the messages matters under different treatment designs. The two separate DID models are as follows:

$$Y_{it} = \alpha_i + \lambda_t + \gamma Post2_t + \beta TwiceQt_i + \delta(TwiceQt_i \times Post2_t) + \varepsilon_{it}, \tag{3}$$

$$Y_{it} = \alpha_i + \lambda_t + \gamma Post2_t + \beta TwiceAvg_i + \delta(TwiceAvg_i \times Post2_t) + \varepsilon_{it}, \tag{4}$$

where $TwiceQt_i$ equals 1 if the entity i gets the quintile information twice, and 0 if the entity i gets the quintile information once. $TwiceAvg_i$ equals 1 if the entity i gets the average information twice, and 0 if the entity i gets the average information once. $Post2_t$ equals 1, representing Period II, and 0 is Period I. The coefficient on the interaction term $TwiceQt_i \times Post2_t$ captures the differential treatment effect between the first-time and the second-time quintile message treatment, and the coefficient on the interaction term $TwiceAvg_i \times Post2_t$ captures the differential treatment effect between the first-time and the second-time mean comparison message treatment.

4 Empirical Results

4.1 Overall Treatment Effects

This research examines whether providing usage feedback motivates residents to change their electricity usage behavior. We randomize different message types, the number of feedback messages sent, and treatment effects among different target groups. Table 3 summarizes the difference-in-differences regression results using equation 2 to equation 4 and presents each sample's mean usage in the "Mean" row. The top half of Table 3 presents the estimation results ($\hat{\delta_1}$ and $\hat{\delta_2}$) of equation 2. The estimated treatment effects of equation 3 and equation 4 are reported in the bottom left and the bottom right of Table 3. Panel A presents the results of the quintile message treatment compared with the control group, and Panel B summarizes the effects of the mean comparison message. We provide the estimation results with the full sample in column (1) in Table 3. These results reveal that neither the quintile message treatment nor the mean comparison message significantly affects residents' electricity behavior no matter how many times the feedback was sent. These results are

similar to those in Anderson et al. (2017), which found that normative messaging has no significant effect on energy conservation in the short run.

One possible explanation for the non-effect is that the social comparison based home energy report's (HER) effectiveness is context-dependent. The electricity usage in this study is lower than the consumption level at studied dormitories in the United States. This argument was posed in Andor et al. (2020), which stated that the cost-effectiveness potentials of HER would be smaller in countries with lower electricity consumption levels. As shown in Table 1, the average electricity usages during our experimental period are between 2.227 - 6.079 kWh per person per week, transforming to 8.908 - 24.316 kWh per month. In comparison, Delmas and Lessem (2014) found that the average electricity usage in UCLA dorms is 198 kWh per person per month. The main source of energy consumption in our experiment in Taiwan comes from cooling, and heating and other appliance usage may be equivalently important in other countries. These differences indicate that the treatment effects of any feedback policy may depend largely on context.

4.2 Heterogeneous Treatment Effects

We next investigate whether treatment effects vary by particular groups of residents. First, we focus on the residents with different levels of electricity use. We restrict the sample to the highest 20% and the lowest 20% users based on the baseline electricity consumption. The results in column (2) and column (3) in Table 3 show that a particular type of message is effective after providing residents the feedback twice. Among the top 20% of users, providing the quintile message twice motivates them to reduce their weekly electricity usage by 0.887 kWh as compared to the first feedback. This effect is sizable, corresponding to 7.66% of their mean usage in Period I and 15.13% of the mean usage of the whole sample. In contrast, the quintile message is not effective among the lowest 20% of users, but the average-type feedback has a significant treatment effect. Providing the lowest 20% of users with the average-type feedback twice associated with usage decreases by 0.201 kWh per person per week (although it is at a 0.1 significance level). This reduction is around 10.99% lower than the mean usage of the lowest 20% of electricity users in Period I.

The heterogeneous treatment effects arise in groups with different electricity usage levels, and the efficacy of the treatment relies on the number of times feedback was provided. The quintile information only affects residents with higher baseline electricity usage, while residents with lower baseline electricity usage are only affected by the mean comparison information. A possible ex-

planation is that the quintile message provides more precise position information than does the average-type information. For the top 20% users, a more accurate message would alert them to their exceptionally high usage among their peers. As a comparison, the quintile information may not work well among the lowest 20% of users as this message may result in the boomerang effect, by which knowing their low usage position induces people to consume more (Schultz et al., 2007). Providing the average-type information to the lowest 20% of users may therefore be more effective as it nudges the users to save energy without triggering the boomerang effect.

We also examine heterogeneous treatment effects by other individual-specific characteristics: students' gender and degree. Tables 4 and 5 summarize the subsample estimation results using equation 2 to equation 4 by students' gender and degree, respectively. According to Table 4, no significant difference is found for the treatment effects between female and male students. Yet, the treatment effects vary by students' degree. The results presented in Table 5 reveal that undergraduate students reduce their electricity consumption by 0.268 kWh per person per week in response to the second-time quintile-type information, which corresponds to a 4.61% lowering of their mean usage in Period I. The nudge message, however, does not have a significant influence on masters and Ph.D. students. We suspect that these results are driven by an economic factor and time spent in the dormitory. Because undergraduate students are less likely to receive scholarships or teaching assistantships than graduate students, they may be more responsive to electricity usage information given its financial pertinence. Undergraduate students, additionally, spend more time in the dormitory than graduate students. We have some evident statistics through our online survey (see Figure 5, Figure 6, and Table 6) - undergraduate students stay 0.3 more hours per day in the dorm than graduate students. The caveat is that the online survey is voluntary, and may be subject to selection bias. Our heterogeneous results suggest that the effectiveness of feedback messages may depend upon the characteristics of the target group.

Apart from the heterogeneous effect, the subsample estimation represented in Table 3 and Table 5 shows that the treatment should be implemented at least twice to drive significant behavioral change. This finding aligns with Fischer (2008), which finds that one-shot information is not enough to encourage conservation behaviors. The result appears to be consistent with the salience theory — the purpose of sending social comparison messages is not only to simply provide relevant information but also to bring people's attention to their electricity consumption.

4.3 Cost-Effectiveness of the Treatment

Table 7 presents a cost-effectiveness analysis. According to our heterogeneity analyses (see Table 3, Table 5, and Table 4), the message treatment is effective among the highest and the lowest 20% baseline electricity users and among undergraduate residents after we provide the usage information twice. Focusing on the highest and the lowest users, we learn that providing the quintile comparison feedback twice to the highest 20% of users (252 residents in our treatment group) saves 0.887 kWh per person per week. Providing the average-type feedback twice to the lowest 20% users saves 0.201 kWh per person per week. Targeting the highest and the lowest 20% users in total saved an estimated 545.94 kWh during our study period (around one month).[8]

The intervention's main cost is the amount paid to collect electricity records of each room. We hired eight part-time students with a wage rate of NT$750 (US$24.25) per session (roughly three hours) to collect electricity records four times since NTHU does not have a smart meter.[9] This wage rate is NT$250 (US$8.08) per hour, which is above the market hourly minimum wage, NT$160 (US$5.17). The electricity feedback messages were delivered by email without extra cost, meaning the total amount spent on the experiment is NT$24,000 (US$776.07).[10] Combining the estimated treatment effect with experiment costs, the cost per 1 kWh reduction in electricity usage is NT$43.96 (US$1.42).[11] On the other hand, providing the second-time quintile comparison message to undergraduate residents (906 residents in our treatment group) saves 485.62 kWh electricity usage in total. The cost-effectiveness of targeting undergraduate residents is, therefore, NT$49.42 (US$1.60) per kilowatt-hour saved.[12]

If the nudging policy is promoted to the whole NTHU residents and we set the wage rate to the market hourly minimum wage, this policy will be far more cost-effective. Bringing this policy into practice, the implementation cost would be NT$11,520 (US$372.51).[13] According to our results,

[8] 545.94 kWh=(0.887kWh)*(252 residents)*(2 weeks)+(0.201kWh)*(246 residents)*(2 weeks).

[9] The exchange rate is 1 United States Dollar to 30.925 New Taiwan Dollar, according to the average exchange rate in 2019.

[10] NT$24,000=(8 part-time students)*(NT$250 hourly wage)*(3 hours per time)*(4 times). Besides the cost of collecting electricity records, we also conduct two online questionnaires to ensure receivers have read the content of the email. Each survey gave away twenty-five NT$100 (US$3.23) Family Mart Gift Vouchers for some randomly drawn twenty-five respondents. This lottery costs NT$5,000 (US$161.68). But, bringing this nudging policy into practice in the future does not require online questionnaires. So we do not count the Gift Vouchers into our cost.

[11] NT$43.96=NT$24,000/545.94kWh.

[12] 485.62 kWh==(0.268kWh)*(906 residents)*(2 weeks). NT$49.42=NT$24,000/485.62kWh.

[13] The last electricity record is only to evaluate the treatment effect. Thus, implementing the policy only needs to collect the electricity record three times. NT$11,520= (8 part-time students)*(NT$160 hourly wage)*(3 hours per time)*(3 times).

if policymakers choose to target the highest and the lowest 20% baseline electricity users (1,345 residents), they would save 2,927 kWh usage during our study period.[14] The cost of saving 1 kWh would be NT$3.936 (US$0.127) in this scenario (see Panel B in Table 7). This number is comparable to the findings of another related study. Chen et al. (2021), which conducted a group contest at another major university in Taiwan. Their results show that providing both inter-ranking and intra-ranking information to all contestants costs NT$8 (US$0.259) to induce a reduction of 1 kWh. Furthermore, if our nudging policy is promoted to universities with a smart meter, the cost of saving 1 kWh of electricity usage would be much less than NT$3.936 (US$0.127).

5 Robustness Checks

One potential source of bias in this study is if treated residents do not pay attention to the electricity feedback messages sent via email (the non-compliance problem). To address this issue, we included a survey link with each electricity email report. By responding to the survey, treated residents confirm that they have read the emailed information, and we can use the survey responses to conduct robustness checks using an Instrumental Variable (IV) framework. In this framework, the original randomization assignment serves as the instrument, and we assume that the randomized messages's effects on treated residents is experienced solely through their reading of the email (the "exclusion restriction" assumption). In this section, we use the survey response to perform a robustness check.

To ensure the robustness of our empirical findings, we estimate the local average treatment effect (LATE) in the IV framework. We use the random assignment to treatment as IVs for respondents (compliers). To estimate the first treatment effect, in the first stage, we run a panel regression model to estimate:

$$\begin{aligned} QtRespondent_{it} &= \theta(QtTreat_i \times Post1_t) + \alpha_i + \lambda_t + \varepsilon_{it}, \\ AvgRespondent_{it} &= \theta(AvgTreat_i \times Post1_t) + \alpha_i + \lambda_t + \varepsilon_{it}, \end{aligned} \tag{5}$$

where $QtRespondent_{it}$ equals 1 if the entity i is assigned to the $QtTreat$ group and has responded to the first survey at time t, and 0 otherwise. $AvgRespondent_{it}$ equals 1 if the entity i is assigned to the $AvgTreat$ group and has responded to the first survey at time t, and 0 otherwise. In the second stage, $QtTreat_i \times Post1_t$ and $AvgTreat_i \times Post1_t$ in equation 2 are being substituted for

[14] 2,972 kWh=(0.887kWh)*(1,345 residents)*(2 weeks)+(0.201kWh)*(1,345 residents)*(2 weeks).

$\widehat{QtRespondent}_{it}$ and $\widehat{AvgRespondent}_{it}$ from the estimation of equation 5. We obtain the LATE estimates from the second stage, i.e., $\widehat{\delta}_{1IV}$ and $\widehat{\delta}_{2IV}$.

Similarly, we use the random assignment $TwiceQt_i \times Post2_t$ and $TwiceAvg_i \times Post2_t$ as instruments for $TwiceQtRespondent_{it}$ and $TwiceAvgRespondent_{it}$ to estimate the second treatment effect. $TwiceQtRespondent_{it}$ equals 1 if respondent i receives the quintile information twice and has responded to the second survey at time t, and 0 if respondent i receives the quintile information once. $TwiceAvgRespondent_{it}$ equals 1 if respondent i receives the average information twice and has responded to the second survey at time t, and 0 if respondent i receives the average information once. We substitute $\widehat{TwiceQtRespondent}_{it}$ and $\widehat{TwiceAvgRespondent}_{it}$ from the estimation of equation 6 to fit equation 3 and equation 4 and obtain $\widehat{\delta}_{IV}$.

$$\begin{aligned} TwiceQtRespondent_{it} &= \theta(TwiceQt_i \times Post2_t) + \alpha_i + \lambda_t + \varepsilon_{it}, \\ TwiceAvgRespondent_{it} &= \theta(TwiceAvg_i \times Post2_t) + \alpha_i + \lambda_t + \varepsilon_{it}, \end{aligned} \tag{6}$$

The estimated LATEs in the IV frameworks are present in Table 8, Table 9, and Table 10. Our baseline estimates are present in Table 3, Table 5, and Table 6. It can be observed that all of the significant baseline estimates remain statistically significant under the IV frameworks, but the magnitude of the LATEs is much larger than the baseline treatment effects. For example, as shown in column (2) of Table 8, providing the quintile message twice to the top 20% of users who respond to surveys motivated them to reduce their weekly electricity usage by 5.589 kWh, compared to the first feedback, which is about six times larger than the baseline estimate of 0.887 kWh in Table 3. Similarly, the estimated LATEs for the lowest 20% users and undergraduate students are statistically significant and have larger magnitudes than the baseline treatment effects.

The results of robustness checks indicate that our estimates are consistent in the IV framework. However, we observe some differences between the OLS and the IV results. Specifically, the LATEs, which represent the average treatment effect among a subgroup of individuals who are most likely to comply with the treatment (compliers), are larger in magnitude than our ITT estimate. (Angrist and Pischke, 2009; Oreopoulos, 2006). This may be because the LATEs are based on a smaller and selected group of individuals compared with the ITT estimate. The ITT adjusting the effects of the non-compliers as follows:

$$LATE = \frac{ITT}{\%\ compliers} \tag{7}$$

In this study, we observed a compliance rate of 15.87% among the top 20% of users in the treatment group that received the quintile message twice. The estimated LATE for this group is

-5.589 kWh, which is approaching the baseline estimate (-0.887 kWh) divided by the proportion of compliers (15.87%). In the treatment group that received the average message twice, the compliance rate among the lowest 20% of users was 32.11%, and the estimated LATE was -0.627 kWh, which is also approaching the baseline estimate (-0.201 kWh) divided by the proportion of compliers (32.11%). Our larger IV results are supported by the fact that ITT estimates are diluted by the presence of non-compliers. These results are consistent with the expectation that LATE estimates would approach the ITT estimates in the presence of non-compliers.

6 Conclusion

In this study, we conducted an experiment to assess the effectiveness of providing non-price-based conservation feedback to students at NTHU in Taiwan through email messages. The experiment involved sending two types of‘ messages to participants: a quintile comparison (Qt group) and a mean comparison (Avg group). We also varied the frequency of message delivery, sending messages either once or twice. The purpose of this experiment was to examine the effects of the feedback on energy conservation behavior by comparing the electricity consumption of the treatment groups to a control group.

Our results indicate that providing feedback on electricity consumption through email has no significant effect on changing residents' average electricity behavior. When examining the treatment effects by different groups of residents, however, we find that the feedback is effective in encouraging conservation behavior of some groups, but only after being provided twice. After receiving the feedback twice, the highest 20% of electricity users reduce their consumption by 7.66% compared to their average usage in Period I, while the lowest 20% of users reduce their consumption by 10.99% compared to their average usage in Period I. Additionally, we find that the feedback is more effective among undergraduate residents compared to graduate students. A back-of-the-envelope calculation suggests that implementing this nudging policy in conjunction with a smart meter infrastructure could be cost-effective for universities looking to conserve electricity.

This study adds to our understanding of conservation efforts in university settings with results determined by conducting a randomized experiment on an entire university population in a context outside the US and by using refined message framing. For policy purposes, our results suggest that the design of conservation messages should be tailored to specific target groups (e.g. high vs. low

electricity users) and should be delivered multiple times to be effective.

It is important to note, however, that this study is limited by its time frame. The experiment was conducted over a short period due to the semester schedule and electricity usage patterns in Taiwan, and thus we were unable to observe long-term behavior change. Future research could explore the dynamics of behavior change over a longer time period.

References

[1] Allcott, H. (2011). Social Norms and Energy Conservation. *Journal of Public Economics, 95*(9–10), 1082-1095.

[2] 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),3003–37.

[3] Allcott, H. (2015). Site selection bias in program evaluation. *Quarterly Journal of Economics*, 130(3), 1117-1165.

[4] Allcott, H. & Kessler, J. B. (2019). The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons. *American Economic Journal: Applied Economics, 11*(1): 236-76.

[5] Allcott, H. & Knittel, C. (2019). Are Consumers Poorly Informed about Fuel Economy? Evidence from Two Experiments. *American Economic Journal, 11*(1), 1–37.

[6] Allcott, H. & Sweeny, R. (2017). The Role of Sales Agents in Information Disclosure: Evidence from a Field Experiment. *Management Science, 63*, 21–39.

[7] Al-Ubaydli, O., List, J. A., LoRe, D., & Suskind, D. (2017). Scaling for economists: Lessons from the non-adherence problem in the medical literature. *Journal of Economic Perspectives*, 31(4), 125-44.

[8] Akerlof, G. A. (1991). Procrastination and obedience. *American Economic Review*, 81(2), 1-19.

[9] Anderson, K., Song, K., Lee, S., Krupka, E., Lee, H., & Park, M. (2017). Longitudinal Analysis of Normative Energy Use Feedback on Dormitory Occupants. *Applied Energy, 189*, 623-639.

[10] Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

[11] Andor, M. A., Gerster, A., Peters, J., Schmidt, C. M. (2020). Social Norms and Energy Conservation Beyond the US. *Journal of Environmental Economics and Management, 103*, 102351.

[12] Asensio, O. I., & Delmas, M. A. (2015). Nonprice Incentives and Energy Conservation. *Proceedings of the National Academy of Sciences, 112*, 510-515.

[13] Asensio, O. I., & Delmas, M. A. (2016). The Dynamics of Behavior Change: Evidence from Energy Conservation. *Journal of Economic Behavior & Organization, 126*, 196-212.

[14] Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2012). The availability and utilization of 401 (k) loans. *Investigations in the Economics of Aging*, 145-176.

[15] Bhanot, S. P. (2017). Rank and response: A field experiment on peer information and water use behavior. *Journal of Economic Psychology*, 62, 155-172.

[16] Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5), 803-843.

[17] Brewer, R. S., Lee, G. E., & Johnson, P. M. (2011). *The Kukui Cup: A Dorm Energy Competition Focused on Sustainable Behavior Change and Energy Literacy.* 2011 44th Hawaii International Conference on System Sciences, Kauai, HI, 1-10.

[18] Breza, E. (2012). Peer effects and loan repayment: Evidence from the krishna default crisis. Job Market Paper MIT, 1.

[19] Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637-644.

[20] Carrell, S. E., Sacerdote, B. I., & West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, 81(3), 855-882.

[21] Chen, C. W., Chen, J.I., & Lin. M. J. (2021). Performance Feedback in a Group Contest: A Field Experiment on Electricity Conservation. *Journal of Economics & Management Strategy*, 1-25.

[22] Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, 56(3), 920-80.

[23] Cialdini, R., Kallgren, C., & Reno, R. R. (1991). A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior. *Advances in Experimental Social Psychology, 24*, 201-234.

[24] Costa, D. L. & Kahn, M. E. (2013). Energy Conservation 'Nudges' And Environmentalist Ideology: Evidence From A Randomized Residential Electricity Field Experiment. *Journal of the European Economic Association, European Economic Association, 11*(3), 680-702.

[25] Dehejia, R., Pop-Eleches, C., & Samii, C. (2021). From local to global: External validity in a fertility natural experiment. *Journal of Business and Economic Statistics*, 39(1), 217-243.

[26] Delmas, M. A., & Lessem, N. (2014). Saving Power to Conserve Your Reputation? The Effectiveness of Private Versus Public Information. *Journal of Environmental Economics and Management, 67*(3), 353-370.

[27] Dixon, G. N., Deline, M. B., McComas, K., Chambliss, L., & Hoffmann, M. (2015). Using Comparative Feedback to Influence Workplace Energy Conservation: A Case Study of a University Campaign. *Environment and Behavior, 47*(6), 667–693.

[28] Eisenberg, D., Golberstein, E., & Whitlock, J. L. (2014). Peer effects on risky behaviors: New evidence from college roommate assignments. *Journal of Health Economics*, 33, 126-138.

[29] Ferraro, P. J. & Price, M. K. (2013). Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment. *The Review of Economics and Statistics, MIT Press, 95*(1), 64-73.

[30] Fischer, C. (2008). Feedback on Household Electricity Consumption: a Tool for Saving Energy? *Energy Efficiency, 1*(1), 79-104.

[31] Gerber, A. S., & Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, 71(1), 178-191.

[32] Ito, K., Ida, T., & Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy*, 10(1), 240-67.

[33] Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2), 263-292.

[34] Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science*, 62(12), 3393-3411.

[35] Kuhnen, C. M., & Tymula, A. (2012). Feedback, self-esteem, and performance in organizations. *Management Science*, 58(1), 94-113.

[36] Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. European Economic Review, 53(1), 1-18.

[37] Meub, L., Runst, P. & von der Leyen, K. (2019). *Can APPealing and More Informative Bills 'Nudge' Individuals into Conserving Electricity*? Center for European, Governance and Economic Development Research Discussion Papers 372, University of Goettingen, Department of Economics.

[38] Myers, E. & Souza, M.(2020). Social Comparison Nudges without Monetary Incentives: Evidence from Home Energy Reports. *Journal of Environmental Economics and Management, 101*, 102315, ISSN 0095-0696.

[39] Nakajima, R. (2007). Measuring peer effects on youth smoking behaviour. *The Review of Economic Studies*, 74(3), 897-935.

[40] Newell, R.G. & Siikamäki, J. (2014). Nudging Energy Efficiency Behavior: The Role of Information Labels. *Journal of the Association of Environmental and Resource Economists, 1*(4), 555-598.

[41] Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.

[42] Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review, 96*(1), 152-175.

[43] Peters, J., Langbein, J., & Roberts, G. (2018). Generalization in the tropics–development policy, randomized controlled trials, and external validity. *The World Bank Research Observer*, 33(1), 34-64.

[44] Schultz, P.W., Nolan, J.M., Cialdini, R.B., Goldstein, N.J., & Griskevicius, V. (2007). The Constructive, Destructive and Reconstructive Power of Social Norms. *Psychological Science, 18*(5), 429–434.

[45] Sunstein, C. R., & Reisch, L. A. (Eds.) (2017). *The Economics of Nudge. Routledge.* Critical Concepts of Economics, Vol.. 4.

[46] Vivalt, E. (2015). Heterogeneous treatment effects in impact evaluation. *American Economic Review*, 105(5), 467-70.

[47] Yakovlev, E. (2018). Demand for alcohol consumption in Russia and its implication for mortality. *American Economic Journal: Applied Economics*, 10(1), 106-49.

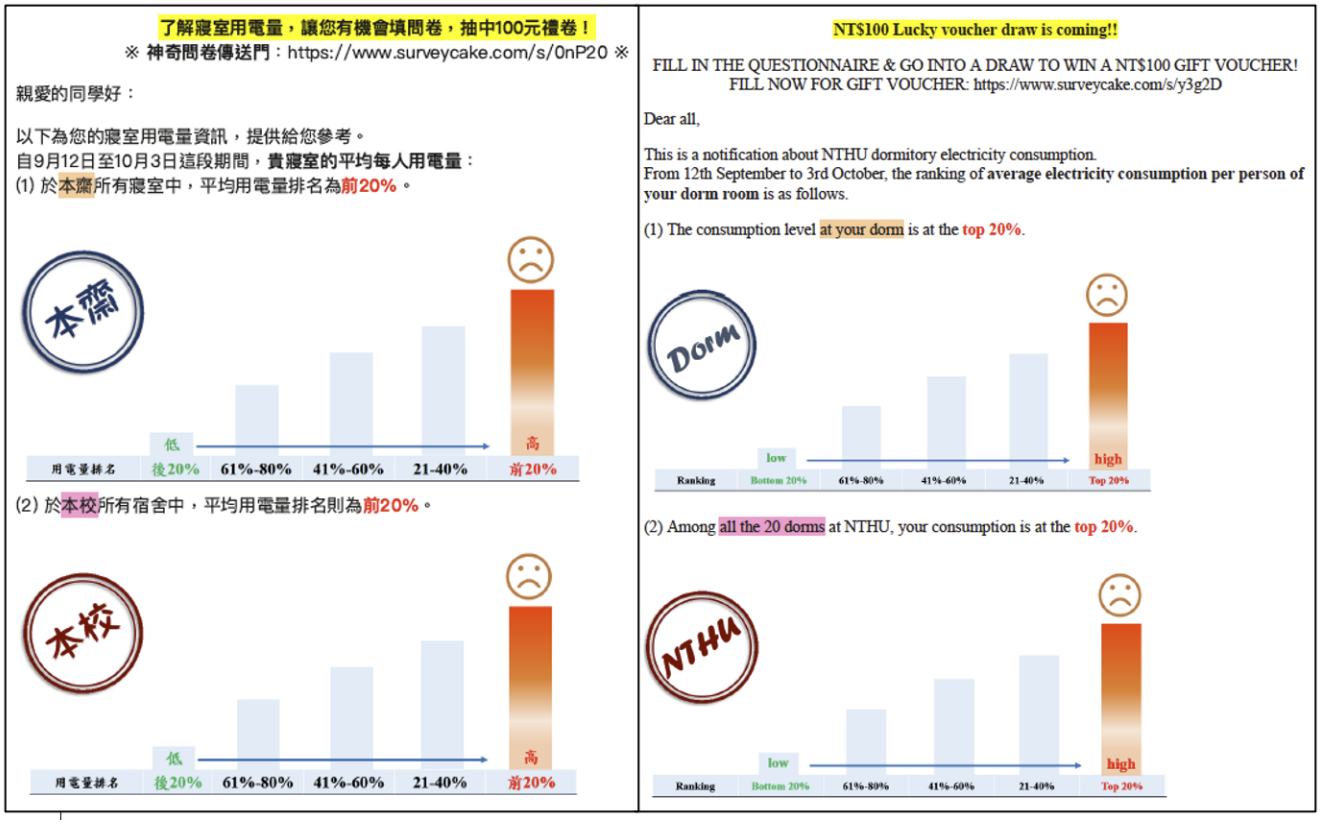

Figure 1: Electricity Feedback – Quintile Information

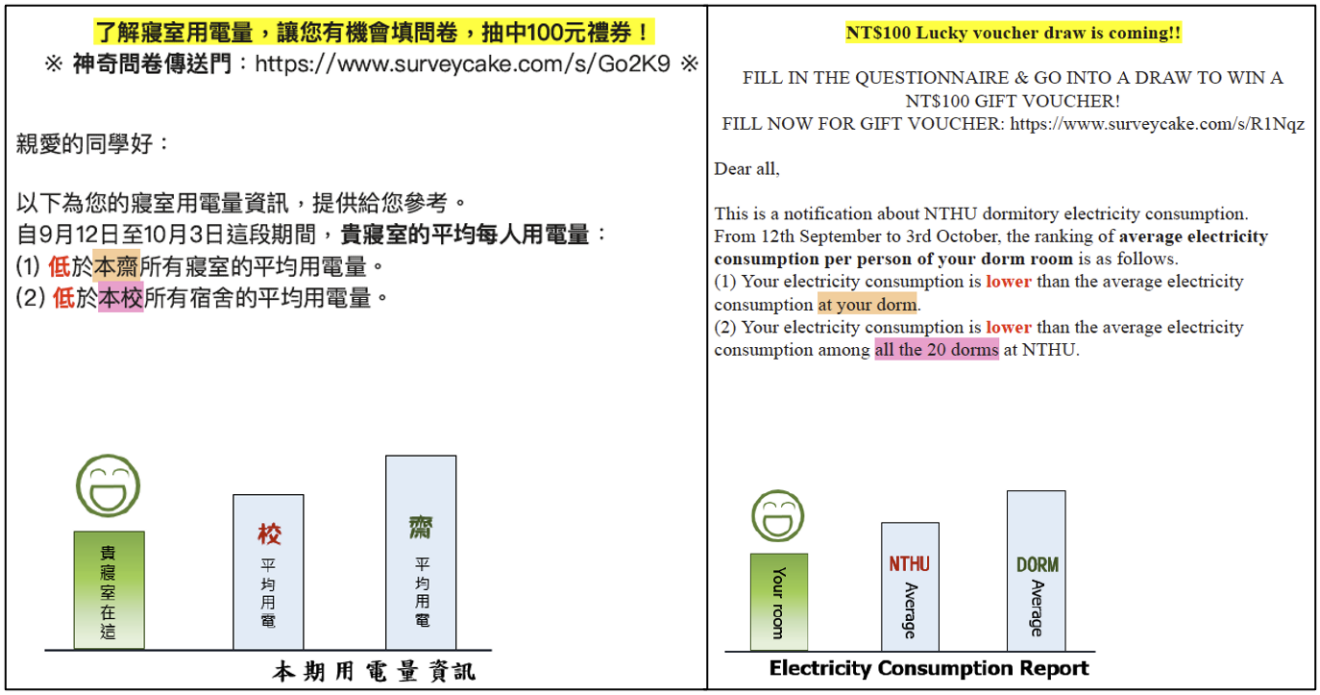

Figure 2: Electricity Feedback – Average Information

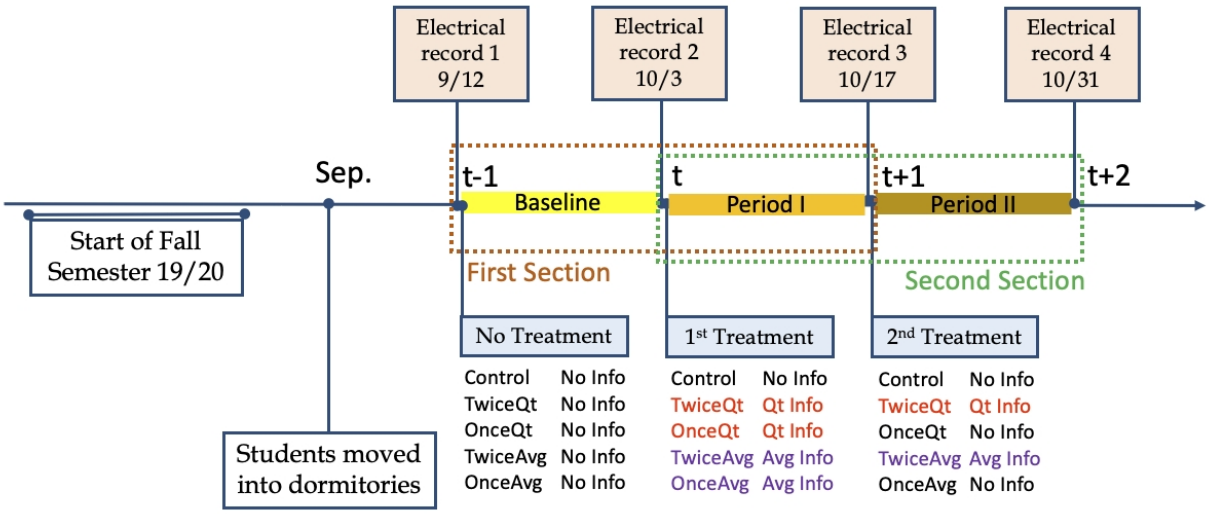

Figure 3: Model Design

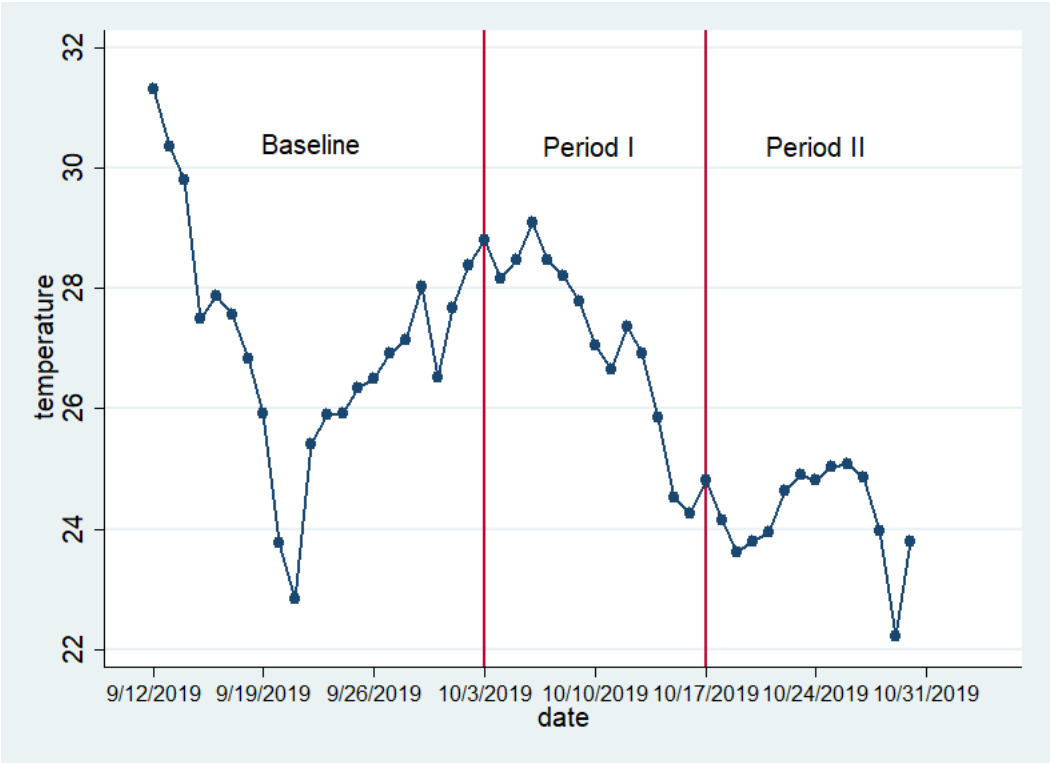

Figure 4: Temperature

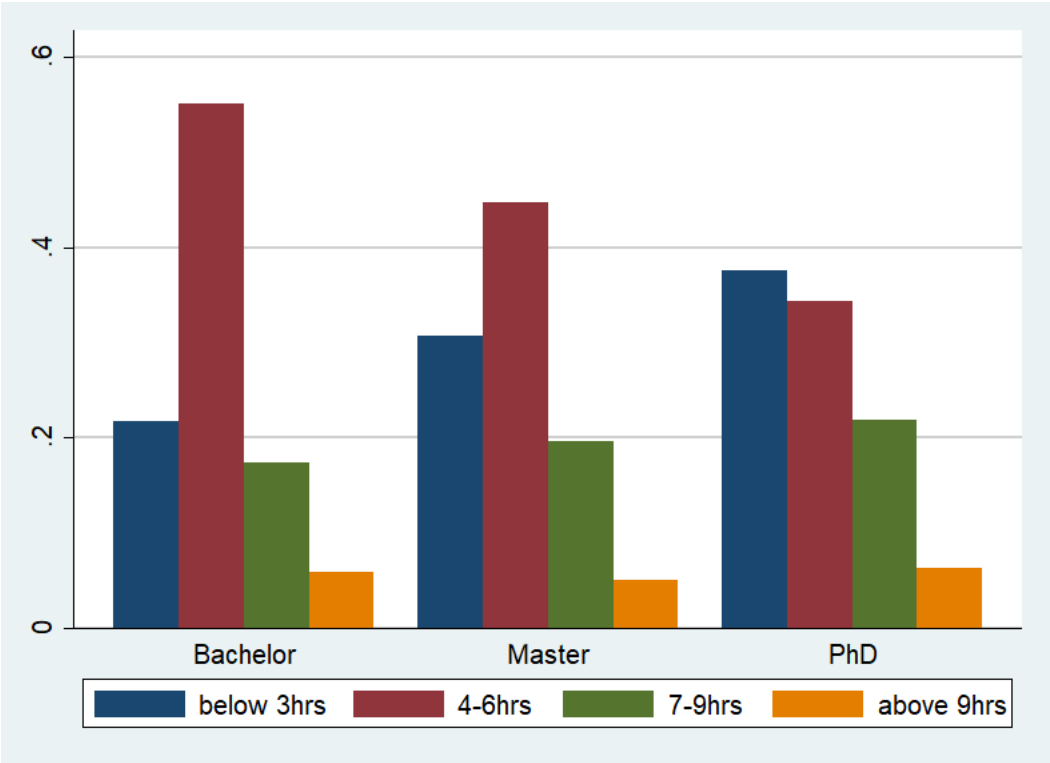

Figure 5: Hours Staying in the Dormitory on Average

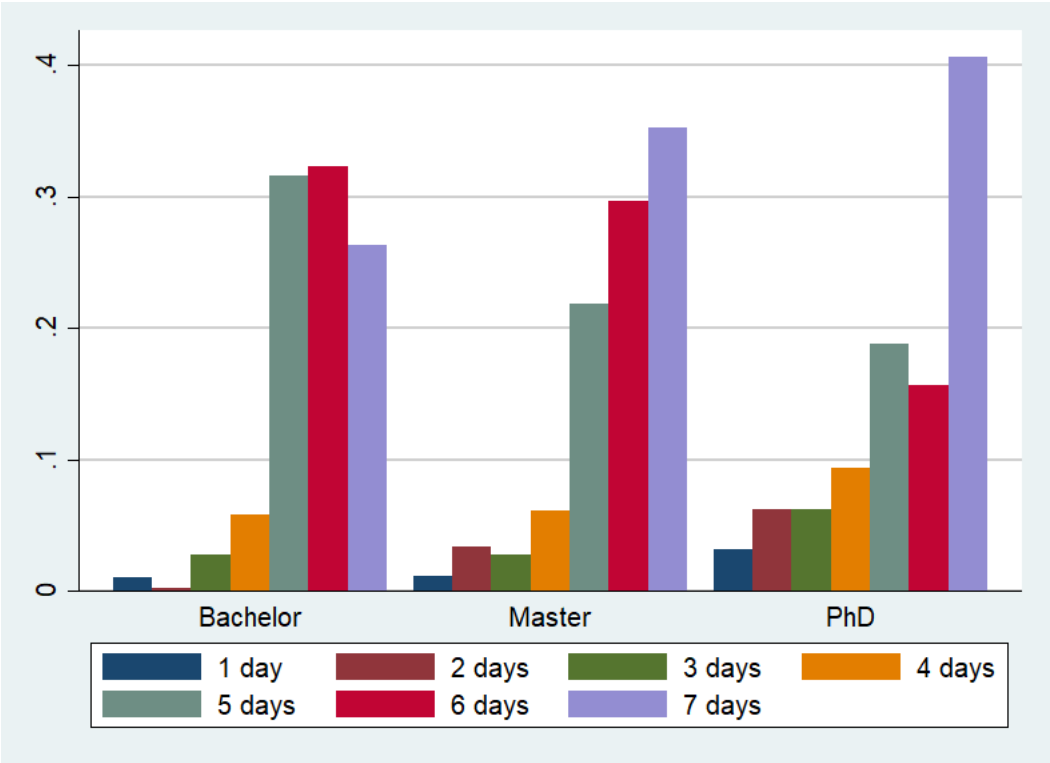

Figure 6: Days Staying in the Dormitory on Average

Table 1: Summary Statistics at the Individual Level by Treatment Status

	All	(1) Control	(2) TwiceQt	(3) OnceQt	(4) TwiceAvg	(5) OnceAvg	Multivariate means test
Variable	N=6,723	N=1,440	N=1,290	N=1,309	N=1,300	N=1,384	Prob>F
Panel A. Electricity Usage							
usage in baseline	5.646	5.697	5.737	5.487	5.783	5.530	0.298
	(4.301)	(4.282)	(4.339)	(4.051)	(4.457)	(4.366)	
usage in period I	6.079	6.179	6.137	5.965	6.119	5.992	0.719
	(4.808)	(4.806)	(4.942)	(4.447)	(4.874)	(4.951)	
usage in period II	2.227	2.347	2.155	2.167	2.325	2.133	0.142
	(2.868)	(3.024)	(2.809)	(2.604)	(3.120)	(2.743)	
Panel B. Characteristics of Rooms							
located floor	3.485	3.466	3.458	3.517	3.477	3.507	0.926
	(1.994)	(1.998)	(1.985)	(2.003)	(1.991)	(1.993)	
# students per room	2.931	2.950	2.930	2.912	2.948	2.915	0.803
	(1.025)	(1.027)	(1.024)	(1.012)	(1.024)	(1.038)	
Panel C. Residents' characteristics							
male	0.582	0.580	0.583	0.581	0.592	0.573	0.899
	(0.493)	(0.494)	(0.493)	(0.494)	(0.492)	(0.495)	
bachelor	0.703	0.722	0.702	0.691	0.694	0.704	0.401
	(0.457)	(0.448)	(0.457)	(0.462)	(0.461)	(0.457)	
master	0.243	0.235	0.242	0.250	0.253	0.236	0.748
	(0.429)	(0.424)	(0.428)	(0.433)	(0.435)	(0.425)	
phd	0.053	0.042	0.055	0.059	0.052	0.059	0.215
	(0.224)	(0.200)	(0.228)	(0.235)	(0.221)	(0.236)	
college	4.319	4.316	4.299	4.300	4.372	4.307	0.932
	(2.379)	(2.381)	(2.387)	(2.420)	(2.420)	(2.293)	

Notes: The usage variables report weekly usage (kWh) per person. Standard deviations are shown in parentheses. Located building ID is the building ID of twenty dormitories. The columns (1), (2), (3), (4) and (5) display the sample means for the five groups. The multivariate means test tests whether the means across our control and four treatment groups are the same.

Table 2: Summary Statistics of Two Online Surveys

	Full Sample	First Survey	Second Survey
Panel A. Objects and Sample			
objects	N=6,723	N=5,283	N=2,590
response sample		N=1,486	N=600
response rate		28.13%	23.17%
error detecting rate		3.8%	13.5%
Panel B. Respondents' Characteristics			
male	0.582	0.489	0.475
	(0.493)	(0.500)	(0.500)
bachelor	0.703	0.688	0.655
	(0.457)	(0.463)	(0.476)
master	0.243	0.264	0.293
	(0.429)	(0.441)	(0.456)
phd	0.053	0.046	0.050
	(0.224)	(0.210)	(0.218)
Panel C. Survey Questions			
intention to reduce usage		0.409	0.393
		(0.492)	(0.489)
care global warm		0.861	
		(0.346)	
hours in the room			5.012
			(2.609)
days in the room			5.685
			(1.232)

Notes: Standard deviations are shown in parentheses. Error detecting rate is a proportion of respondents whose answer of their usage ranking is inconsistent with our record to the whole respondents. "Intention to reduce usage" and "care global warm" are dummy variables, 1 for yes and 0 for not. "Hours in the room" is the average hours that they spend in the dormitory per day after deducting sleep time. "Days in the room" is the average days that they spend in the dormitory per week.

Table 3: Impact of Quintile and Average Information on Usage by Electricity Usage

	Info: Quintile				Info: Average		
	(1)	(2)	(3)		(1)	(2)	(3)
	Whole	Highest 20%	Lowest 20%		Whole	Highest 20%	Lowest 20%
QtTreat×Post1	-0.043	-0.081	-0.137	AvgTreat×Post1	-0.082	0.013	-0.086
	(0.068)	(0.201)	(0.106)		(0.068)	(0.214)	(0.099)
Mean	5.863	11.585	1.829	Mean	5.863	11.585	1.829
N	13,446	2,544	2,596	N	13,446	2,544	2,596
TwiceQt×Post2	-0.184	-0.887**	-0.079	TwiceAvg×Post2	0.065	0.583	-0.201*
	(0.141)	(0.440)	(0.143)		(0.138)	(0.438)	(0.116)
Mean	4.106	8.082	1.211	Mean	4.140	8.590	1.485
N	5,198	992	978	N	5,368	1,008	1,052
Time fixed effects	Yes	Yes	Yes	Time fixed effects	Yes	Yes	Yes
Entity fixed effects	Yes	Yes	Yes	Entity fixed effects	Yes	Yes	Yes

Notes: The outcome variable is the usage (kWh) per person per week during each period. Standard errors in parentheses are clustered at the individual level. * $p < 0.1$, **$p < 0.05$, ***$p<0.01$. Residents are classified into "Highest 20%" and "Lowest 20%" subsample based on their levels of electricity usage in the Baseline. QtTreat includes the OnceQt group and TwiceQt group, the quintile comparison information group. AvgTreat includes the OnceAvg group and TwiceAvg group, the average information group. QtTreat×Post1 and AvgTreat×Post1 report the treatment effects from equation 2. TwiceQt×Post2 and TwiceAvg×Post2 report the treatment effects from equation 3 and equation 4 respectively. "Mean" refers to the mean usage of each sample during each period.

Table 4: Impact of Quintile and Average Information on Usage by Students' Gender

	Info: Quantile			Info: Average	
	(1)	(2)		(1)	(2)
	Female	Male		Female	Male
QtTreat×Post1	-0.038	-0.047	AvgTreat×Post1	-0.124	-0.050
	(0.106)	(0.089)		(0.105)	(0.089)
Mean	4.985	6.495	Mean	4.985	6.495
N	5626	7820	N	5626	7820
TwiceQt×Post2	-0.289	-0.110	TwiceAvg×Post2	0.127	0.004
	(0.240)	(0.171)		(0.234)	(0.165)
Mean	3.078	4.845	Mean	3.318	4.729
N	2174	3024	N	2242	3126
Time fixed effects	Yes	Yes	Time fixed effects	Yes	Yes
Entity fixed effects	Yes	Yes	Entity fixed effects	Yes	Yes

Notes: The outcome variable is the usage (kWh) per person per week during each period. Standard errors in parentheses are clustered at the individual level. * $p < 0.1$, **$p < 0.05$, ***$p<0.01$. QtTreat includes the OnceQt group and TwiceQt group, the quintile comparison information group. AvgTreat includes the OnceAvg group and TwiceAvg group, the average information group. QtTreat×Post1 and AvgTreat×Post1 report the treatment effects from equation 2. TwiceQt×Post2 and TwiceAvg×Post2 report the treatment effects from equation 3 and equation 4 respectively. "Mean" refers to the mean usage of each sample during each period.

Table 5: Impact of Quintile and Average Information on Usage by Students' Degree

Info: Quintile				Info: Average			
	(1)	(2)	(3)		(1)	(2)	(3)
	Bachelor	Master	PhD		Bachelor	Master	PhD
QtTreat×Post1	-0.121	0.157	0.134	AvgTreat×Post1	-0.093	-0.025	-0.082
	(0.077)	(0.149)	(0.398)		(0.076)	(0.150)	(0.406)
Mean	5.810	6.047	5.731	Mean	5.810	6.047	5.731
N	9,452	3,266	714	N	9,452	3,266	714
TwiceQt×Post2	-0.268*	0.149	-0.619	TwiceAvg×Post2	0.085	0.020	0.157
	(0.161)	(0.313)	(0.695)		(0.153)	(0.315)	(0.711)
Mean	4.169	4.024	3.731	Mean	4.103	4.367	3.576
N	3,620	1,278	296	N	3,752	1,312	298
Time fixed effects	Yes	Yes	Yes	Time fixed effects	Yes	Yes	Yes
Entity fixed effects	Yes	Yes	Yes	Entity fixed effects	Yes	Yes	Yes

Notes: The outcome variable is the usage (kWh) per person per week during each period. Standard errors in parentheses are clustered at the individual level. * $p < 0.1$, **$p < 0.05$, ***$p<0.01$. QtTreat includes the OnceQt group and TwiceQt group, the quintile comparison information group. AvgTreat includes the OnceAvg group and TwiceAvg group, the average information group. QtTreat×Post1 and AvgTreat×Post1 report the treatment effects from equation 2. TwiceQt×Post2 and TwiceAvg×Post2 report the treatment effects from equation 3 and equation 4 respectively. "Mean" refers to the mean usage of each sample during each period.

Table 6: Hours and Days Staying in the Dormitory on Average by Degrees of Programs

	Undergraduates	Graduates	
	Mean	Mean	Diff.
times in the room	5.111	4.799	-0.313*
days in the room	5.687	5.692	0.005
	N=396	N=211	

Table 7: Cost Effectiveness

Grouping Basis	Usage		Degree
Target Group	Highest 20%	Lowest 20%	Bachelor
In our experiment			
Information type	Quintile	Average	Quintile
Second TE (kwh/pre week)	-0.887	-0.201	-0.268
Treated residents	252	246	906
Treated weeks	2	2	2
Total saving kWh	-447.048	-98.892	-485.616
Cost (NT$)	24,000		24,000
Cost-Effectiveness (NT$/kwh)	43.961		49.422
Apply it to whole target sample			
Whole residents	1,345	1,345	4,726
Treated weeks	2	2	2
Total saving kWh	-2386.03	-540.69	-2533.136
Cost (NT$)	11,520		11,520
Cost-Effectiveness (NT$/kwh)	3.936		4.548

Notes: The exchange rate is 1 United States Dollar to 30.925 New Taiwan Dollar, the average exchange rate in 2019.

Table 8: Robustness checks: Impact of Quintile and Average Information on Usage by Electricity Usage

	Info: Quintile				Info: Average		
	(1)	(2)	(3)		(1)	(2)	(3)
	Whole	Highest 20%	Lowest 20%		Whole	Highest 20%	Lowest 20%
$\widehat{QtRespondent}$	-0.159	-0.450	-0.324	$\widehat{AvgRespondent}$	-0.282	0.059	-0.228
	(0.251)	(1.117)	(0.253)		(0.234)	(1.010)	(0.262)
N	13,446	2,544	2,596	N	13,446	2,544	2,596
$\widehat{TwiceQtRespondent}$	-0.837	-5.589*	-0.260	$\widehat{TwiceAvgRespondent}$	0.267	3.905	-0.627*
	(0.647)	(2.954)	(0.469)		(0.567)	(3.006)	(0.363)
N	5,198	992	978	N	5,368	1,008	1,052
Time fixed effects	Yes	Yes	Yes	Time fixed effects	Yes	Yes	Yes
Entity fixed effects	Yes	Yes	Yes	Entity fixed effects	Yes	Yes	Yes

Notes: The outcome variable is the usage (kWh) per person per week during each period. Standard errors in parentheses are clustered at the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p<0.01$. Residents are classified into "Highest 20%" and "Lowest 20%" subsample based on their levels of electricity usage in the Baseline. *QtRespondent* equals 1 if the entity is assigned to the *Qt* group and has responded to the first survey. *AvgRespondent* equals 1 if the entity is assigned to the *Avg* group and has responded to the first survey at time *t*. *TwiceQtRespondent* equals 1 if the entity gets the quintile information twice and has responded to the second survey, and 0 if the entity gets the quintile information once. *TwiceAvgRespondent* equals 1 if the entity gets the average information twice and has responded to the second survey, and 0 if the entity *i* gets the average information once. $\widehat{QtRespondent}$, $\widehat{AvgRespondent}$, $\widehat{TwiceQtRespondent}$, and $\widehat{TwiceAvgRespondent}$ report the LATEs in the IV framework.

Table 9: Robustness checks: Impact of Quintile and Average Information on Usage by Students' Gender

Info: Quantile			Info: Average		
	(1)	(2)		(1)	(2)
	Female	Male		Female	Male
$\widehat{QtRespondent}$	-0.115	-0.203	$\widehat{AvgRespondent}$	-0.346	-0.210
	(0.321)	(0.386)		(0.292)	(0.370)
N	5626	7820	N	5626	7820
$\widehat{TwiceQtRespondent}$	-1.072	-0.595	$\widehat{TwiceAvgRespondent}$	0.398	0.023
	(0.903)	(0.930)		(0.733)	(0.872)
N	2174	3024	N	2242	3126
Time fixed effects	Yes	Yes	Time fixed effects	Yes	Yes
Entity fixed effects	Yes	Yes	Entity fixed effects	Yes	Yes

Notes: The outcome variable is the usage (kWh) per person per week during each period. Standard errors in parentheses are clustered at the individual level. * $p < 0.1$, **$p < 0.05$, ***$p<0.01$. *QtRespondent* equals 1 if the entity is assigned to the *Qt* group and has responded to the first survey. *AvgRespondent* equals 1 if the entity is assigned to the *Avg* group and has responded to the first survey at time *t*. *TwiceQtRespondent* equals 1 if the entity gets the quintile information twice and has responded to the second survey, and 0 if the entity gets the quintile information once. *TwiceAvgRespondent* equals 1 if the entity gets the average information twice and has responded to the second survey, and 0 if the entity *i* gets the average information once. $\widehat{QtRespondent}$, $\widehat{AvgRespondent}$, $\widehat{TwiceQtRespondent}$, and $\widehat{TwiceAvgRespondent}$ report the LATEs in the IV framework.

Table 10: Robustness checks: Impact of Quintile and Average Information on Usage by Students' Degree

Info: Quintile				Info: Average			
	(1)	(2)	(3)		(1)	(2)	(3)
	Bachelor	Master	PhD		Bachelor	Master	PhD
$\widehat{QtRespondent}$	-0.466	0.501	0.566	$\widehat{AvgRespondent}$	-0.317	-0.087	-0.361
	(0.298)	(0.477)	(1.679)		(0.259)	(0.513)	(1.778)
N	9,452	3,266	714	N	9,452	3,266	714
$\widehat{TwiceQtRespondent}$	-1.252*	0.589	-3.999	$\widehat{TwiceAvgRespondent}$	0.383	0.067	0.553
	(0.761)	(1.233)	(4.699)		(0.692)	(1.078)	(2.511)
N	3,620	1,278	296	N	3,752	1,312	298
Time fixed effects	Yes	Yes	Yes	Time fixed effects	Yes	Yes	Yes
Entity fixed effects	Yes	Yes	Yes	Entity fixed effects	Yes	Yes	Yes

Notes: The outcome variable is the usage (kWh) per person per week during each period. Standard errors in parentheses are clustered at the individual level. * $p < 0.1$, **$p < 0.05$, ***$p<0.01$. *QtRespondent* equals 1 if the entity is assigned to the *Qt* group and has responded to the first survey. *AvgRespondent* equals 1 if the entity is assigned to the *Avg* group and has responded to the first survey at time *t*. *TwiceQtRespondent* equals 1 if the entity gets the quintile information twice and has responded to the second survey, and 0 if the entity gets the quintile information once. *TwiceAvgRespondent* equals 1 if the entity gets the average information twice and has responded to the second survey, and 0 if the entity *i* gets the average information once. $\widehat{QtRespondent}$, $\widehat{AvgRespondent}$, $\widehat{TwiceQtRespondent}$, and $\widehat{TwiceAvgRespondent}$ report the LATEs in the IV framework.

Appendix A Online Questionnaires

Q1. Your electricity consumption is

(quintile group)

a. top 20% (higher usage)

b. 21-40%

c. 41-60%

d. 61-80%

e. bottom 20% (lower usage)

(average group)

a. higher than the average dorm consumption

b. lower than the average dorm consumption,

Q2. When you received the electricity consumption message, you will

a. try to reduce the usage

b. live as usual

Q3. Do you care about the issue about climate change or global warming in daily life?

a. Of course! I am very concerned about this kind of information!

b. Oh, I am not very concerned about such issues.

Q4. How long do you spend in the dormitory on average (after deducting sleep time)?

a. less than 3 hours

b. 4 to 6 hours

c. 7 to 9 hours

d. More than 9 hours

Q5. How many days do you spend in the dormitory per week on average?

a. 1 day; b. 2 days; c. 3 days; d. 4 days; e. 5 days; f. 6 days; g. 7 days