Performance Feedback in a Group Contest: A Field Experiment on Electricity Conservation*

Chia-Wen Chen† Josie I Chen‡ Min-Jen Lin§

January 2021

Abstract

We conduct a field experiment to study whether revealing both the competitive state and the social state in a group contest affects individual beliefs and efforts. Our experiment randomizes group composition, participation in the contest, and types of information received in the contest. We find that contestants without feedback about relative performance had difficulty assessing their group's competitive status, and laggards within a group tended to be overconfident about their relative contribution. In addition, we find that contestants receiving both competitive and social information were more likely to have correct beliefs about their positions during the contest and exerted the most effort, while contestants receiving no performance feedback did not behave differently from those who did not participate in the contest. Our results support the notion that providing feedback matters in a group contest.

Keywords: Relative Performance Feedback; Field Experiment; Group Contest; Energy Conservation.

---

*We are grateful to David Rapson, Joseph Tao-yi Wang, Yi-Yi Chen, and seminar participants at the 2017 European ESA Meeting, National Tsing Hua University, 2018 Contests: Theory and Evidence Conference, 2018 Taiwan Economic Research for helpful comments. Liang-Cheng Chen, Jiang-Shiang Hu, Chi-Bo Lin, and Hsiang-En Ho provided excellent research assistance. We thank the Ministry of Science and Technology (grants 107-2410-H-001-024-SS2 and 109-2636-H-002-007) for financial support.

†Contact information: Research Center for Humanities and Social Sciences, Academia Sinica, e-mail: cwzchen@gate.sinica.edu.tw.

‡Corresponding author. Contact information: Department of Economics, National Taiwan University, e-mail: josiechen@ntu.edu.tw.

§Contact information: Department of Economics, National Taiwan University, e-mail: mjlin@ntu.edu.tw.

1 Introduction

Group contests are common in daily life. Prizes and honors are often awarded to groups with top performances. Firms battle to win standards wars or patent races; hospitals, charity organizations, and universities compete to be listed as the best organizations; and within firms, sales teams often compete in tournaments for rewards. A group contest inevitably creates two different states within each group member during the contest. The first is how good the group is compared to other groups (the competitive state), and the second is how good the individual's performance is compared to other group members (the social state).[1] In this paper, we conduct a field experiment to examine the impacts of revealing either or both of these states on efforts exerted in such contests.

A growing body of literature supports that a group contest is an effective tool for encouraging individual efforts.[2] Erev, Bornstein, and Galili (1993) and Nalbantian and Schotter (1997) demonstrated that adding a group-level competition component in a public goods setting increases cooperation within groups.[3] While studies have shown that individuals care about how their group performance compares with those of other groups (Tan and Bolle, 2007; Janssen et al., 2016), it is well established in various settings that individuals also care about their individual performances in comparison with those of other individuals (Frey and Meier, 2004; Croson and Shang, 2008; Chen, Harper, Konstan, and Li, 2010; Allcott, 2011). [4] A less explored issue, however, is how individuals react to their own relative performance

[1] We follow previous studies in the literature in applying the term 'social' to emphasize the impact of social comparison when one receives intra-group information feedback, which is totally independent of monetary payoff (Frey and Meier, 2004; Croson and Shang, 2008; Chen, Harper, Konstan, and Li, 2010; Allcott, 2011). Notably, as we will see in the next section, there are *no* social interactions among subjects in our experiment.

[2] In general, there are two types of contest: individual and group. We expect group contests yield better performances than individual contests because contestants in a group contest are expected to have at least the additional motivation of guilt aversion. For example, Chen and Lim (2013) directly compared the motivations in group and individual contests. They showed that guilt aversion made contestants more willing to invest efforts when they were in group contests.

[3] Regarding group coordination problems, Bornstein, Gneezy, and Nagel (2002) also showed that inter-group competition improves coordination and cooperation within groups.

[4] Tan and Bolle (2007) found that providing competition information alone is able to boost the contribution in a public goods setting. Janssen, Lee, and Sundaram (2016) extended this study in a large online experiment under the framing of sustainable development and found that feedback in a competitive state is able to raise

within a group, in addition to information about their group's relative performance during a group contest. This issue thus motivates our research question: What type of information about relative performance encourages the most efforts in a group contest? On the one hand, when competitive information is provided, a competitive effect exists and works through an increase in testosterone. Mazur and Booth (1998) showed that people are more motivated to compete after winning a competition because testosterone levels are enhanced in winners. On the other hand, as Festinger (1954) highlighted, people rely on social information in order to fulfill the drive to evaluate themselves. People are motivated to work harder in order to perform as well as their peers. Most of all, if competitive and social information are not substitutes, we expect that providing both types of ranking information will motivate people more than providing only a single type of information.

To understand the joint effects of competitive and social information, we perform a randomized control trial (RCT) that assigns different information treatments to participants in a group contest. We test whether revealing information about relative performance across teams (competitive information) and/or within teams (social information) has any effect on beliefs and efforts exerted in the group contest. The empirical setting is an electricity-saving team contest in residence halls at a major university in Taiwan. During the 5-week contest, we sent weekly feedback (via emails and text messages) to contest participants, and we randomly assigned treatment structure, receiving inter- or intra-team ranking information—or both—at the team level. We exploit a panel dataset of electricity usage at the room-hour level, covering time periods before and after the contest, to estimate the treatment effects. The data structure allowed us to include a set of hourly and room-by-day-of-week-level fixed effects to account for heterogeneity in usage across time periods and rooms.

We find that beliefs about competitive and social states in the contest affected contestants' efforts during the contest. Firstly, contestants without ranking information were uncertain about their states during the contest. Laggards on the team tended to be overcon-

eco-friendly movement by approximately 5%.

fident about their relative contribution within the team, and contestants tended to report that their team was an average team, regardless of the group's actual competitive position. Secondly, we find that contestants receiving *both* inter- and intra-team rankings reduced electricity consumption the most, while contestants receiving no ranking information about their relative performance reduced electricity consumption the least. Lastly, evidence from the endline survey supported that participants receiving both types of information were most likely to find information affecting their energy use. These results suggested that contestants in the group contest cared about their competitive states and social states, but without our intervention, they could not identify their exact states. Our results support that providing both competitive and social information feedback in a group contest is an effective tool to elicit greater effort.

Our paper builds on the existing literature on the effects of group incentives and information on individual behaviors.[5] In public good games, Erev et al. (1993) and Nalbantian and Schotter (1997) showed that competition between groups alleviate free-riding within that group. Group incentives also promote productivity (Hamilton, Nickerson, and Owan, 2003; Bandiera, Barankay, and Rasul, 2013; Blimpo, 2014; Babcock, Bedard, Charness, Hartman, and Royer, 2015), prosocial lending, and charitable giving behavior, possibly due to social identity, goal-setting, or the reluctance to 'let down one's team' (Chen, Chen, Liu, and Mei, 2017; Charness and Holder, 2019).[6]

Empirical work by Allcott (2011), Ayres et al. (2013), Costa and Kahn (2013), Delmas and Lessem (2014), and Kandul, Lang, and Lanz (2020) supports that social comparison information is effective in reducing residential electricity usage.[7] Discussing the intra-team

[5]For studies on the effects of social information or peer effects on productivity, see Falk and Ichino (2006), Mas and Moretti (2009), Bandiera et al. (2010), Kuhnen and Tymula (2012), Azmat and Iriberri (2016), Azmat et al. (2019), Barankay (2011), Barankay (2012), and Gill, Kissová, Lee, and Prowse (2019).

[6]Chen, Chen, Liu, and Mei (2017) showed that lenders in a crowd-lending community contribute more when they join teams. Charness and Holder (2019) found that team incentives with competition led to more donations than individual incentives with competition.

[7]In a study on residential water demand, Ferraro and Price (2013) also found that providing social comparison messages had a large influence on consumption.

and inter-team information in a public goods setting similar to ours, Augenblick and Cunha (2015) showed that donors behave differently when receiving information about past contribution behaviors from the competing party versus from the same party. However, our paper differs from the literature described by studying the effect of providing *both* social and competitive information in a group contest.[8] We find that receiving information about both competitive and social states was the most effective way to reduce electricity consumption during the contest, indicating that the effects of competitive and social information are not simply substitutes.

The remainder of this paper is organized as follows. Section 2 describes the experimental design, Section 3 discusses the data and the empirical setting, Section 4 quantifies the results of the experiment, and Section 5 concludes.

2 Experimental Design

The experiment was conducted during the 2017 spring semester at a major university in Taiwan. Three dormitories were located on the main campus of the university, and each dormitory was equipped with smart meters that recorded room-level electricity usage every 15 minutes. Electricity bills were sent to individual students after each semester—usually along with their tuition fees for the next semester—so students had no information regarding their electricity usage during the semester. At the beginning of the 2016 academic year, which lasted from September 2016 to June 2017, there were 1054 residents.[9]

We sent email invitations to all undergraduate students living in the dormitory halls to complete a baseline survey about their dormitory life. We recruited 553 respondents from 216 rooms to complete the baseline survey. The sample is a good representation of the

[8]In a laboratory setting, Böhm and Rockenbach (2013) and Cárdenas and Mantilla (2015) also provided inter-group and intra-group information to participants. However, Böhm and Rockenbach (2013) conducted the experiment in a competition-free set-up, and Cárdenas and Mantilla (2015) focused on the treatment effect of introducing competition into a public goods setting, whereas our paper studies the treatment effects of receiving different types of information about relative performance.

[9]Undergraduate students made up 89% of that number, and 84% of those were first-year students. The number of dorm rooms is limited, and so first-year students have priority for residentship.

undergraduate residents at the university.[10] We anticipated sufficiency of power to test for between-treatment differences—see Appendix A for more detail.

We randomly constructed teams to participate in our electricity-saving contest, making sure that before the contest, each team was considered 'average.' This ensured the contest outcome would be uncertain, and each team had a fair chance to win. In doing so, each team consisted of a low-user room, an average-user room, and a heavy-user room.[11] We randomized the treatment status at the team level and the treatment status of each participant varied according to whether they were assigned to be a contestant as well as the type of information they received during the contest. We randomly assigned teams into one control group that did not compete in the contest and received contest-irrelevant text messages, and four treatment groups that competed in the contest. Each treatment group was given different information treatment. The no-ranking treatment group received contest-irrelevant text messages similar to the control group; the inter-ranking treatment group received information about rankings between teams (leading, middle, or behind); the intra-ranking treatment group received information about rankings within their own team (leading, middle, or behind); and finally, the both-ranking treatment group received both inter- and intra-team ranking information. The number of rooms in the control group was 45. The number of rooms receiving no-ranking, inter-ranking, intra-ranking, and both-ranking information was 45, 42, 42, and 42, respectively. Appendix Table A1 provides the experimental design and the number of recruited participants in each treatment group.

We informed all participants assigned to the treatment groups (contestants) that they had been chosen to participate in an electricity-saving contest. We then invited them to read the competition rules and take a quick test to ensure their comprehension. According to the

[10] The distribution of the respondents for all colleges is: College of Law (13%), College of Business (29%), College of Public Affairs (27%), College of Social Sciences (15%), College of Humanities (11%), and College of Electrical Engineering and Computer Science (5%), which is close to distribution for the school's entire population, which is 15%, 28%, 23%, 15%, 11%, and 8%, respectively.

[11] We first calculated the average usage per person in each room using data from April 2017 (one month prior to recruitment) and split all 216 rooms evenly into three types: heavy user, average user, and low user. Next, we randomly drew three rooms from the three types—one for each type—to construct a team.

competition rules, all competing dorm rooms would be divided into teams—each consisting of 3 rooms—and teams would be constructed so that every room had a chance to win. The top one-third of teams using the least electricity per person, during the competition period would win, receiving a small reward shared equally among team members. The expected amount of the reward payment would be NT$150 (around US$5) per person.[12] To control for the endowment effect, study participants in the control group received a lottery ticket with a prize value of NT$150, and they were told that winning tickets (one-third of all tickets) would be drawn after the study ended. Contestants were also informed that member identities across rooms would be kept anonymous. There was no interaction between team members.

The 5-week electricity-saving contest was held from May 10, 2017 to June 11, 2017. The competition period was from 7:00 a.m. to 10:00 p.m. each day during this time period.[13] Figure 1 shows the timeline of the experiment. At 10:00 p.m. each Sunday, participants received text messages based on their treatment status—contest-irrelevant or based on their relative performance in the contest.[14] Each participant received four messages during the contest period, and the cutoff of each 'week' in our study was 10 p.m. on Sunday. In particular, during the first week of the contest, none of the participants received any contest-relevant ranking information, so we expected behavioral responses resulting from treatments to emerge only after the first week.

The contest ended at 10:00 p.m. on June 11, 2017. After that, we asked all study participants to complete an endline survey before June 14, 2017. Out of 553 participants,

[12]Each contest participant belonging to a winning team would receive NT$100 for sure and a chance to win other prizes with an expected value of NT$50.

[13]Because the peak demand for electricity in Taiwan was during daytime hours rather than at nighttime, we restricted the contest to be from 7:00 a.m. to 10:00 p.m. to make the study more policy-relevant.

[14]All 553 participants knew they were participating in a study about their dormitory life that required them to answer short quizzes every Sunday night based on text messages they received. However, they did not have information about what types of messages they would receive. Study participants were also provided with a small monetary incentive—NT$20, or US$0.67—to complete a quiz right after receiving their messages.

399 completed the endline survey.[15] On June 18, 2017, we informed the control group and treatment group participants of the lottery and contest results and gave them corresponding payments. The semester ended on June 25, 2017.[16] Supplementary material from the experiment can be found in Appendix A.

3 Data and Empirical Set-up

Two types of appliances were used to generate electricity-usage data in the dorm rooms: regular appliances—such as lights, computers, hairdryers, etc.—subject to 110-volt outlets, and air conditioners, which were subject to 220-volt outlets. An air conditioner (AC) was in each room, but there were no other large appliances in the dorm rooms; cooking and using any other large appliances in the rooms was strictly prohibited due to safety reasons. A typical air conditioner was a 1625-watt model, meaning that it consumes 1.625 kWh if it runs nonstop for an hour. A back-of-the-envelope calculation suggested that the energy consumed by using the AC for a full hour could support lighting the room for 10.2 hours, a computer for 10.8 hours, a fan for 32.5 hours, and a hairdryer for 1.1 hours.[17] As a result, we expected that the decision to use the AC would drive the results of the electricity-saving contest. Since most people in Taiwan are advised by the utility company to set their AC temperatures between 26°C and 28°C, we expected to find less use of AC on the experiment when temperatures were lower than 28°C.[18]

We collected the readings for the 110- and 220-volt outlets from the university's dormitory electricity usage database. This data allowed us to calculate real-time total usage and AC usage for each individual room at the hourly level from April 2017 to June 2017. Figure 2 presents the usage distribution by room prior to recruitment. On average, low-user, average-

[15] Although not all of the participants answered the endline survey, we note that our attrition rate is zero—none of the participants dropped out of our experiment during the contest.

[16] Students moved out of their dorm rooms after the semester, and so we cannot measure the persistence of our treatments after the semester.

[17] We assume that lighting requires four 40-watt fluorescent tubes and that a computer(with monitor), a fan, and a hair dryer use 150 watts, 50 watts, and 1500 watts of electricity, respectively.

[18] The state-owned Taiwan Power Company has monopoly over the electricity retail market.

user, and heavy-user rooms consumed 5.6, 9.2, and 16.4 kWh per capita prior to recruitment. In terms of usage per room, even though heavy-user rooms consist of one-third of the total rooms, they consumed about 50% of the total electricity during this time.

We also collected hourly temperatures and daily maximum and minimum temperatures at the weather station located closest to the university. Figure 3 shows hourly temperatures from April 1, 2017 to June 25, 2017. As shown in the figure, average temperatures were rising during the spring semester. Nevertheless, there were several cool days that did not require the use of air conditioning. We addressed this issue by defining time periods when we were much more likely to detect changes in behavior in our empirical analysis. 'Hot days' were defined as having an average temperature of at least 28°C or a maximum daily temperature of at least 30°C. We presented 'hot days' as the shaded area in Figure 3. Those were the days that participants had to make deliberate decisions regarding using the AC in their rooms, and we expected the effect of the experiment to be stronger on those days.

Sometimes the meter readings were not transmitted to the school's database, resulting in missing values for certain time periods and rooms. To make sure that we compared each room's efforts in the contest correctly, we constructed a balanced panel. We ended up having 311 sample hours during the contest, 346 hours before the contest—from April 1 to May 6, before students were invited to the contest—and 112 hours after the contest. Our main results only used data before and during the experiment, and the total number of observation was 141,255 for 215 rooms.[19]

3.1 Summary Statistics

Table 1 provides summary statistics for room-level hourly electricity usage, *before* and *during* the contest, by treatment status. All usage data ranged from 7:00 a.m. to 10:00 p.m., during which timeframe contestants were asked to save electricity. Column (1) gives the

[19] There was one room without usage data, and so we made sure that this room was assigned to the control group after the final randomization process. Therefore, we only have usage data for 44 rooms in the control group, but all participants in the control group were invited to answer the baseline and endline surveys.

means of usage from the control group, including standard deviations shown in parentheses. Columns (2), (3), (4), and (5) provide differences in means between the control group and the groups that received no ranking information, inter-ranking information, intra-ranking information, and both-ranking information, respectively, with the corresponding p-values shown in parentheses. For example, a p-value reported in column (2) is associated with a test for the difference between two means using 89 observations (44 rooms in the control group and 45 rooms in the no-ranking group.)

Panel A of Table 1 provides summary statistics for hourly usage before the contest. We did not find significant (at the 5% level) differences in total or AC usage (220-volt outlets) across treatment status before the contest; however, participants in the control group seemed to use more electricity from 110-volt outlets than those receiving no-ranking information and inter-ranking information. Panel C of Table 1 further provides information about room types across treatment status. The majority of rooms in the sample were four-bed rooms and female-only. To conclude, our randomization process successfully removed differences across groups for our main variables of interest, such as total electricity usage and AC usage. However, it did not completely eliminate differences across groups for certain variables.[20] In our empirical specification, we will explicitly address this challenge.

Panel B of Table 1 provides summary statistics for hourly usage during the contest. During this time period, 67% of electricity consumed by participants in the control group was from using ACs. The mean and standard deviation of usage from 220-volt outlets were 0.223 kWh and 0.164 kWh, respectively. These were large compared to the mean and standard

[20]We find that control-group rooms were more likely to be four-bed rooms compared to the other groups. Appendix Table B1 provides additional summary statistics for several individual-level characteristics. For most of the variables in the baseline survey, we did not find systematic differences across the control and treatment groups. There were no significant differences across groups in terms of effort levels exerted in a room, number of credits in the semester, hours put into part-time jobs, extra curricular activities, the amount of the previous semester's electricity bills, and whether the bills were paid by students' families. However, we do find that compared to participants in the control group, participants receiving no-ranking information were less likely to be in their dorm rooms, which may explain why they tended to use less electricity from 110-volt outlets. In addition, participants receiving no-ranking information and inter-ranking information seemed to have lower levels of satisfaction (with their dorm life) compared to participants in the control group.

deviation from 110-volt outlets, which were 0.108 kWh and 0.039 kWh, respectively. These results suggested that most of the variation within the contest was AC driven. Comparing groups during the contest, we find suggestive evidence showing that providing information seemed to help reduce contestants' usage. Nevertheless, we have to address the issue that some variables at the room level were not balanced, even before the experiment. Therefore, in our empirical setting, we rely on a difference-in-differences approach, which allowed for room-by-day-of-week-level fixed effects to control for permanent differences across rooms.

3.2 Graphical Analysis

Figures 4(a), 4(b), 4(c), and 4(d) plot daily average AC usage per room from 7:00 a.m. to 10:00 p.m. for each treatment group in dashed lines. Because the majority of rooms in our data were four-bed rooms, to make the figures more comparable, we focus on four-bed rooms.[21] In each figure, the usage from the control group is shown as a solid line, and the competition period is shown as the shaded area. Looking at the usage data before the experiment, we find that the control group's consumption typically shared the same fluctuation patterns and trends as each treatment group; as such, the future usage of the control group was very likely a good predictor of each treatment group. On the other hand, usage from the control group seemed to be higher than some treatment groups before the experiment, suggesting that failing to control for permanent differences across rooms would over-estimate savings from the contest.

Comparing usage patterns between Figures 4(a), 4(b), 4(c), and 4(d), we find that the gap between the control group and the treatment group seemed to be the largest for participants receiving both-ranking information treatment and smallest for participants receiving no information treatment. Moreover, it seemed that most of the treatment effects occurred at the end of the contest after contest participants had received the fourth (and final) notification about their performance. This result is consistent with findings in the literature evaluating

[21] In Appendix Figure B1, we normalize consumption per capita and plot the normalized average AC usage including all participating rooms. The results are robust to this specification.

the dynamic effects of nonlinear incentive contracts (Misra and Nair, 2011; Larkin, 2014), in which agents would increase efforts within the time frame most favorable for them to 'close the deal.' Another possible factor contributing to this end-of-contest result is that the temperatures were higher during the final week of the contest, making behavioral responses more likely to be observed. Finally, we find that the usage gap shrank after the experiment, especially for the treatment group receiving both-ranking information.

3.3 Empirical Set-up

Even though each team's composition and treatment status was randomly assigned, balance across treatment groups was not automatically guaranteed, especially when we did not have a large sample. Previous discussion points out that there existed pre-experiment differences in some observed—and probably unobserved—variables across treatment status. To address this challenge, we analyze the data using a 'difference-in-differences' approach under a randomized control trial setup. Let y_{it} denote the outcome variable (such as AC usage) for room i during hour t, and the baseline regression is

$$\begin{aligned}(1) \qquad y_{it} = {} & \alpha_{it} + \alpha_t + \beta_0 1(\text{Week 1})_t \times 1(\text{In contest})_i + \beta_1 1(\text{Info weeks})_t \times 1(\text{No ranking})_i \\ & + \beta_2 1(\text{Info weeks})_t \times 1(\text{Inter ranking})_i + \beta_3 1(\text{Info weeks})_t \times 1(\text{Intra ranking})_i \\ & + \beta_4 1(\text{Info weeks})_t \times 1(\text{Both rankings})_i + \epsilon_{it},\end{aligned}$$

where α_{it} and α_t are room-by-day-of-week and hour-of-sample fixed effects, respectively, and ϵ_{it}, is the idiosyncratic error term. The room-by-day-of-week fixed effects captured permanent unobserved factors that affected the outcome variable at the room level (such as the location of a room), taking into account occupants' schedules throughout the week (Monday, Tuesday, etc.) during the semester, while hour-of-sample fixed effects controlled variation from permanent unobserved factors at the hourly level (such as a special weather condition). The indicator variables $1(\text{In contest})_i$, $1(\text{No ranking})_i$, $1(\text{Inter ranking})_i$, $1(\text{Intra ranking})_i$, and $1(\text{Both rankings})_i$ take a value of 1 if room i received the corresponding treatment and 0 otherwise. In addition, $1(\text{Week } 1)_t$ and $1(\text{Info weeks})_t$ are indicator variables for the first week (during which none of the contestants received any information) and weeks 2–5 (during which information varied by treatment status), respectively. We expect β_k $(k = 0, 1, 2, 3, 4)$ to be negative if our treatments were effective in obtaining greater efforts from contestants compared to those in the control group. Because the assignment of treatment groups was completely random, we are unlikely to encounter an endogeneity problem in this setting. All standard errors are clustered at the team level.

4 Results

In this section, we first discuss results regarding contestants' beliefs about their relative performance at the end of the contest. Then, we provide results regarding contestants' behavioral responses by treatment status.

4.1 Did Contestants Without Ranking Information Have Correct Beliefs About Their Relative Performance?

After the contest, we elicited contestants' beliefs about their *final* intra- and inter-team positions in the endline survey.[22] Table 2 reports belief matrices for intra-team positions, which collect the relationships between actual and self-reported intra-team positions by treatment status. Here, we use a, b, and c to denote the *actual* leading, middle, and behind position *within* a team. Similarly, we use $a^{'}$, $b^{'}$, and $c^{'}$ to denote the *self-reported* leading, middle, and behind position within a team. In Table 2, the column with heading a reports the shares of contestants whose within-team positions are actually a at the end of the contest and who self-reported that their within-team positions are $a^{'}$, $b^{'}$, or $c^{'}$.[23] If all contestants

[22]Contestants were provided with a small monetary incentive (NT$20, or US$0.67) to guess their final positions at the end of the contest correctly.

[23]We construct indicator variables for each self-reported position, calculate their averages within a room, and report room averages in each cell. The results are robust to the use of individual averages instead of room averages.

had correct beliefs about their within-team positions, we expected column a to be $(1, 0, 0)'$, and the belief matrix collecting columns a, b, and c would be an identity matrix.

As shown in Table 2, a majority of contestants receiving our intra-ranking or both-ranking treatments correctly predicted their relative positions, suggesting that our information treatments were effective. In particular, the diagonal elements of the matrix for contestants receiving both-ranking information are 0.97, 0.80, and 0.92. By contrast, contestants receiving no-ranking information could not assess their positions properly. The diagonal elements of the associated belief matrix are 0.63, 0.32, and 0.29. In fact, for contestants receiving no-ranking information, columns a, b, and c in the belief matrix are $(0.63, 0.23, 0.14)$, $(0.30, 0.32, 0.38)$, and $(0.08, 0.63, 0.29)$, respectively, suggesting that low users (position a) tended to know that they used less electricity than others, average users could not tell whether they used more or less than others, and heavy users tended to believe that they were average users. It is also interesting to point out that even though team laggards were over-confident about their contribution within a team, most of them knew that they were not low users. To compare changes in beliefs for cells receiving ranking information, we calculate their differences in beliefs relative to corresponding cells receiving no-ranking information, with the p-values reported in parentheses. We find that providing either intra- or both-ranking information significantly changed participants' beliefs about their position within a team.

Table 3 provides belief matrices for inter-team positions. In this case, we use A, B, and C to denote the leading, middle, and behind position *across* teams, respectively. We find that the majority of contestants without any ranking information reported that they were in the middle position regardless of the team's actual position. Once they were provided feedback about their weekly performance, contestants were more likely to correctly discover their positions at the end of the contest. In particular, the diagonal elements of the belief matrix for the both-ranking treatment are 0.79, 0.64, and 0.80, while those for the no-ranking treatment are 0.35, 0.64, and 0.11.[24]

[24]One may be concerned that intra-team information seemed to help teams in the losing position shape

The elicited beliefs from contestants suggest that contestants were not able to tell whether their teams were winning or losing without information, and both average users and heavy users were more likely to be either uncertain or had biased beliefs about their relative performance within a team. If the nature of competitive states and/or social states in a group were important for contestants to determine their efforts, we expect our information treatments would change their behavior during the contest.

4.2 Did Information Treatment Increase Contestants' Efforts?

Table 4 gives the difference-in-differences estimation results using equation (1). The outcome variables used in each regression are shown in the column headings for columns (1)–(4). We provide the estimation results with the full sample in the first two columns, and results with the restricted sample—hot days—are in the next two columns. All regressions contain room-by-day-of-week and hour-of-sample fixed effects; the estimated coefficients should therefore be interpreted as changes in outcome within a room, controlling for each room's weekly schedule and each hour's weather condition.

Firstly, contestants without ranking information behaved as if they were not competing at all. Across columns (1)–(4), the coefficients associated with the time period without information for all contestants (β_0), or the no-ranking group during the information weeks (β_1), are all small and not statistically significant.[25] By contrast, the estimated coefficients for contestants that received the information treatment are at least two times as large as those for β_1, although not all of these are precisely estimated. Among the estimated coefficients for

beliefs about their inter-team position (37% of rooms guessed correctly that their teams were behind in the contest compared to 11% without ranking information). However, we note that the difference is only statistically significant at the 10% level. We cannot rule out the possibility that in certain rare cases, participants may be able to infer their inter-team position after receiving intra-team information. For example, heavy-user room participants may infer that they were in a 'bad' team when they received intra-team information suggesting that their room was not always behind within the team. Nevertheless, we did not find such evidence in our data.

[25] To interpret our findings, insignificance results from the contest coefficients (without information) may suggest that the contest prize was too small in our context to induce individual effort or that a group contest is ineffective by itself in general. Our interpretation is that even though sometimes the effect of a group contest is insignificant (due to a small monetary reward), providing feedback in such case can still be powerful. We thank an anonymous referee for raising this point.

β_1, β_2, β_3, and β_4, those for β_4—receiving both-ranking information—are the largest and are statistically significant across all columns. For hot days, the estimated coefficients for β_4 are -0.081 kWh in hourly total usage and -0.083 kWh in hourly AC usage, which in magnitude are 81 and 21 times as large as those for β_1. Therefore, we find that providing both inter- and intra-team rankings was the most effective treatment to reduce electricity consumption during the contest. In fact, the p-values associated with the hypothesis that both-ranking information—on top of the effect of competing in a contest—is irrelevant ($\beta_4 = \beta_1$) are 0.073 and 0.084 in the first two columns (all days) and 0.007 and 0.010 in the last two columns (hot days).

To explore heterogeneous effects across competitive and social states, we also estimate the model by splitting samples according to a room's states. For example, if receiving favorable information about the competitive state (such as leading in the contest) helped leading contestants exert more effort, we expect $(\beta_2 - \beta_1)$ to be negative when estimating the model using only samples from those leading the contest at the end of the first week. However, we lack the statistical power to conduct inferences in several cases. We provide our analysis of the heterogeneous effects in Appendix C.

4.3 Evidence of Behavioral Responses from the Endline Survey

Results from previous subsections provide evidence on behavioral responses using actual usage data. In this subsection, we examine participants' responses in the endline survey. One question we asked in the endline survey was: 'Did weekly text messages affect your energy use?' Out of 533 recruits from our experiment, 399 (or 72%) answered the above question. Our findings would be strengthened if results from actual usage data and the endline survey are consistent.

In Table 5, we provide the regression results where the dependent variable is an indicator variable equal to 1 if the answer to the survey question is 'Yes' and 0 otherwise, and the independent variable is treatment status. The base group in the regression is the control group in the experiment. Interestingly, among control group participants who were not in the contest and received completely energy-irrelevant information, 17.8% reflected that weekly text messages affected their energy use. In comparison, contestants in the no-ranking group—who received exactly the same messages as those in the control group—were only 2% more likely than those in the control group to find text messages affecting their energy use, and the difference was not statistically significant. This result again confirms that participants in the control group and no-ranking group indeed received the same information, which was unrelated to their performance in the contest. More importantly, contestants who received inter-ranking, intra-ranking and both-ranking information were 17.8%, 14.3%, and 24% more likely—than those in the control group—to find weekly text messages affecting their energy use, and the estimated differences are all statistically significant at the 10% level. Furthermore, we can reject the hypothesis that each of the corresponding types of information was irrelevant in the contest—compared to the no-ranking group—with p-values of 0.06, 0.165, and 0.01, respectively. The above results show that contestants receiving both-ranking information in the group contest were the most likely to find such information helpful.

4.4 Robustness Checks

In this section, we provided several robustness checks. Table 6 provides estimation results of equation (1) under alternative specifications. First, instead of using room-by-day-of-week fixed effects as in our preferred setting, we explore other room-level fixed effects, such as room and room-by-hour-of-day fixed effects. We also consider room-by-day-of-week-by-time-of-day fixed effects, in which case we allow for individual time-of-day (morning, afternoon, night) fixed effects for each room and for these effects to vary by day of week. Even though the majority of rooms are four-bed rooms, some rooms are smaller and only have 2 or 3 occupants. To ease the concern that coordination may vary by room size, in another specification, we restrict samples to four-bed rooms only. In addition, to reduce the effects of outliers, we estimate the model with a logarithm transformation of usage as our dependent variable. Specifically, we use the inverse hyperbolic sine transformation to allow for zeros in the outcome variable.[26] The estimation results in Table 6 show that our results are robust to alternative econometric specifications and concerns regarding room size.

Our preferred specification is a difference-in-differences setting. As a robustness check, we present the results without using pre-experiment data in Appendix Table B2. The disadvantages of relying only on data during the contest are that we cannot remove permanent differences across rooms and lose precision. In fact, the estimates are the same as those shown in panel B from Table 1, but standard errors (and thus p-values) differ in two tables. We also estimate the model without pre-experiment data under a random-effect model (Table B3). We find our results are robust to these alternative specifications.

4.5 Persistence of the Treatments and Total Savings During the Contest

Up to now, our estimation only uses data before and during the contest. To show how the treatment effects varied during the contest and to test their persistence after the contest, we estimate a model on AC usage similar to that in equation (1) but with additional data *after* the contest. We also include full interaction terms between each treatment group and 7 weeks of data—5 weeks during the contest and 2 weeks after the contest. We estimate the model in one regression and present the estimated coefficients (along with their 95% confidence intervals) from each treatment group in Figure 5. Confirming the patterns shown in Figure 4, the estimated coefficients for those who received ranking information are largest (in terms of magnitude) during the final week of the contest. The treatment effect seemed to diminish after the contest. Because we only had two weeks of data after the contest, we cannot trace long-term behavior responses after the contest. However, unlike the persistent treatment effect shown in some long-run OPOWER programs (Allcott and Rogers, 2014), the post-contest results do not support that the effects of the contest would persist. The difference

[26]The inverse hyperbolic sine (arcsinh) transformation of a variable x is defined as $\text{arcsinh}(x) = \ln(x + \sqrt{x^2+1})$.

might be explained by the short duration of the contest, which was not long enough for participants to form new energy-saving habits. Additionally, restrictions in the dormitories and the small amount of possible reward might have stopped participants from investing in more energy-efficient appliances that could reduce their energy consumption even if their energy use patterns remain unchanged.

We calculate the total electricity saved during the contest and presented the results in Table 7. In column (1), we present hourly savings (kWh/per room) based on the balanced data. On average, the hourly electricity saved by a contest-participating room was 0.032 kWh, which corresponds to a 10.6% saving in electricity per hour, and a total of 1,712 kWh saved throughout the contest. The above estimates are obtained from the balanced panel, which accounts for 311 hours during the contest. However, the actual number of hours during the contest was 465 (15 hours per day times 31 contest days); thus, using estimates from the balanced panel may underestimate the savings from the contest.

To address the problem of missing data at the hourly level, we interpolate hourly usage data. Furthermore, to prevent us from relying too much on hourly data, we calculate each room's daily usage during the contest. In doing so, we find that the daily savings per contest room were 0.425 kWh, which corresponded to a 10.1% reduction in hourly electricity usage, and the total savings in electricity from the contest were 2,251 kWh. To put the above savings into context, using estimates of the short-run elasticity of residential electricity demand in Taiwan (Kuo et al., 2018) and the United States (Reiss and White, 2008), which are -0.653 and -0.1, respectively, our savings can be interpreted as a price increase ranging from 15% to 101%.[27]

Regarding external validity, Allcott and Mullainathan (2010) examined the effect of usage reports sent by the OPOWER Company to millions of households in the United States, finding that behavioral interventions reduce electricity consumption by 2%. Delmas, Fis-

[27] Although a US$5 prize may seem small and insufficient to stimulate energy conservation action, competition and information itself may make energy use more salient as well as introduce a moral imperative to conserve. We thank an anonymous referee for suggesting this point.

chlein, and Asensio (2013) conducted a meta-analysis of experimental studies showing that information-based energy conservation treatments on average reduce usage by 7.4%. Unlike typical households in previous studies, our participants were not directly responsible for their utility bills.[28] This factor could explain why treatments had stronger behavioral responses in our study than in previous work, in which many low-cost energy saving options may have already been explored even without interventions.

The total costs of the contest were NT$25,374 (NT$24,259 from prize and NT$1,115 from sending text messages). Thus, in the current setup, it cost NT$11 to induce a reduction of 1 kWh. If, instead, both-ranking information was given to all contestants, a back-of-the-envelope calculation suggests that the cost to induce a reduction of 1 kWh would be NT$8; this was less than the amount (NT$10) in the demand response program in Taiwan wherein the utility company paid industrial users to cut back their electricity consumption during demand peaks. A final point is that sending text messages to contestants accounted for about 4% of the contest's budget. Nevertheless, our findings suggest that most of the savings from the contest came from contestants receiving messages about their ranking-information during the contest.

5 Concluding Remarks

Group incentives can be a powerful tool to help individuals make productive decisions (Charness and Sutter, 2012). We add empirical evidence to the literature of group incentives by estimating the effect of receiving information about relative performance on beliefs and efforts exerted during a group contest. We show that without intervention, some contestants had biased beliefs about their intra-team positions and were uncertain about how well their teams competed in the contest. In addition, contestants receiving information about their competitive and social states increased their efforts. Our results suggest that feedback about

[28] We note that 89% of participants reported in the baseline survey that their tuition and fees (including electricity bills) are paid by their parents.

relative performance matters in a group contest.

Our experimental setting and results have several implications for future researchers or practitioners conducting group contests to promote individual efforts. Firstly, our results suggest that group contests *per se* do not necessarily lead to higher efforts. While several studies in the literature have shown that group contests can lead to higher individual efforts, we do not find such an effect for contestants without ranking information in our field setting. In the worst-case scenario, we would have conducted the same group contest but omitted any interim information feedback. In this case, we would still have sent prizes to 'winners' in the contest, but these contestants may not have saved much electricity at all. Secondly, providing interim performance feedback can be a cost-effective way to induce effort from contestants. The algorithm we used to calculate interim performance feedback was almost the same as the algorithm used to calculate the final outcome of the contest; processing outcomes during the contest therefore incurred few additional costs. Moreover, sending text messages to contestants only accounted for about 4% of our contest's budget. Nevertheless, most of the energy savings from the contest came from contestants receiving feedback about their relative performance.

Our study has several limitations and can be extended in a number of directions. Firstly, the sample size of the experiment was limited by the number of dorm rooms in our setting. Secondly, we randomized each participating room's treatment status after teams were randomly constructed but before the contest. Therefore, our information treatments were applied to contestants irrespective of their competitive or social states during the contest. While our design gave us enough variation to identify the average treatment effects, the design was not tailored to study the heterogeneous effect of information at particular states, such as the effect of information for contestants who were winning in the contest and falling behind within a team. We explored heterogeneous effects by conditioning contestants' usage at the end of the first week, but many of the coefficients of heterogeneous effects could not be precisely estimated due to the limitation of our sample size and design. Future research could address these issues by increasing the sample size or randomizing the treatment status at states during the contest to explore the full pattern of the effects of performance feedback.

We also note that our study was not suitable for examining the heterogeneous effect of information by group size because the majority of rooms in our data are four-bed rooms. More research is needed on the extent to which information treatments address the coordination problem in a team contest and how the treatment effects vary by group size in a field setting. Finally, member identity was kept anonymous in our experiment. While communications between team members may have helped them learn how to better save electricity or coordinate their behavior, this channel was shut down in our setting. Incorporating other mechanisms to help group members make decisions as well as investigating the effect of information feedback in other settings are promising areas for future research.

References

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics 95*(9), 1082–1095.

Allcott, H. and S. Mullainathan (2010). Behavior and energy policy. *Science 327*(5970), 1204–1205.

Allcott, H. and T. Rogers (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review 104*(10), 3003–37.

Augenblick, N. and J. M. Cunha (2015). Competition and cooperation in a public goods game: A field experiment. *Economic Inquiry 53*(1), 574–588.

Ayres, I., S. Raseman, and A. Shih (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *The Journal of Law, Economics, and Organization 29*(5), 992–1022.

Azmat, G., M. Bagues, A. Cabrales, and N. Iriberri (2019). What you don't know...can't hurt you? A natural field experiment on relative performance feedback in higher education. *Management Science 65*(8), 3714–3736.

Azmat, G. and N. Iriberri (2016). The provision of relative performance feedback: An analysis of performance and satisfaction. *Journal of Economics & Management Strategy 25*(1), 77–110.

Babcock, P., K. Bedard, G. Charness, J. Hartman, and H. Royer (2015). Letting down the team? Social effects of team incentives. *Journal of the European Economic Association 13*(5), 841–870.

Bandiera, O., I. Barankay, and I. Rasul (2010). Social Incentives in the Workplace. *The Review of Economic Studies 77*(2), 417–458.

Bandiera, O., I. Barankay, and I. Rasul (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association 11*(5), 1079–1114.

Barankay, I. (2011). Rankings and social tournaments: Evidence from a crowd-sourcing experiment. Working paper, University of Pennsylvania.

Barankay, I. (2012). Rank incentives: Evidence from a randomized workplace experiment.

Working paper, University of Pennsylvania.

Blimpo, M. P. (2014). Team incentives for education in developing countries: A randomized field experiment in Benin. *American Economic Journal: Applied Economics 6*(4), 90–109.

Böhm, R. and B. Rockenbach (2013). The inter-group comparison—intra-group cooperation hypothesis: Comparisons between groups increase efficiency in public goods provision. *PLOS ONE 8*(2), 1–7.

Bornstein, G., U. Gneezy, and R. Nagel (2002). The effect of intergroup competition on group coordination: An experimental study. *Games and Economic Behavior 41*(1), 1–25.

Cárdenas, J. C. and C. Mantilla (2015). Between-group competition, intra-group cooperation and relative performance. *Frontiers in Behavioral Neuroscience 9*, 33.

Charness, G. and P. Holder (2019). Charity in the laboratory: Matching, competition, and group identity. *Management Science 65*(3), 1398–1407.

Charness, G. and M. Sutter (2012). Groups make better self-interested decisions. *Journal of Economic Perspectives 26*(3), 157–76.

Chen, H. and N. Lim (2013). Should managers use team-based contests? *Management Science 59*(12), 2823–2836.

Chen, R., Y. Chen, Y. Liu, and Q. Mei (2017). Does team competition increase pro-social lending? Evidence from online microfinance. *Games and Economic Behavior 101*, 311–333.

Chen, Y., F. M. Harper, J. Konstan, and S. X. Li (2010). Social comparisons and contributions to online communities: A field experiment on MovieLens. *American Economic Review 100*(4), 1358–98.

Costa, D. L. and M. E. Kahn (2013). Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association 11*(3), 680–702.

Croson, R. and J. Y. Shang (2008). The impact of downward social information on contribution decisions. *Experimental Economics 11*(3), 221–233.

Delmas, M. A., M. Fischlein, and O. I. Asensio (2013). Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. *Energy*

Policy 61, 729–739.

Delmas, M. A. and N. Lessem (2014). Saving power to conserve your reputation? The effectiveness of private versus public information. *Journal of Environmental Economics and Management 67*(3), 353–370.

Erev, I., G. Bornstein, and R. Galili (1993). Constructive intergroup competition as a solution to the free rider problem: A field experiment. *Journal of Experimental Social Psychology 29*(6), 463–478.

Falk, A. and A. Ichino (2006). Clean evidence on peer effects. *Journal of Labor Economics 24*(1), 39–57.

Ferraro, P. J. and M. K. Price (2013). Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *The Review of Economics and Statistics 95*(1), 64–73.

Festinger, L. (1954). A theory of social comparison processes. *Human relations 7*(2), 117–140.

Frey, B. S. and S. Meier (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *American Economic Review 94*(5), 1717–1722.

Gill, D., Z. Kissová, J. Lee, and V. Prowse (2019). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science 65*(2), 494–507.

Hamilton, B. H., J. A. Nickerson, and H. Owan (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy 111*(3), 465–497.

Janssen, M. A., A. Lee, and H. Sundaram (2016). Stimulating contributions to public goods through information feedback: Some experimental results. *PloS ONE 11*(7), 1–16.

Kandul, S., G. Lang, and B. Lanz (2020). Social comparison and energy conservation in a collective action context: A field experiment. *Economics Letters 188*, 108947.

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

Kuo, C.-H., C.-D. Yuan, H.-C. Chai, and F.-K. Ko (2018). The estimation of price elasticities of energy demand in Taiwan. *Journal of Taiwan Energy 5*(1), 27–45.

Larkin, I. (2014). The cost of high-powered incentives: Employee gaming in enterprise software sales. *Journal of Labor Economics 32*(2), 199–227.

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–45.

Mazur, A. and A. Booth (1998). Testosterone and dominance in men. *Behavioral and brain sciences 21*(3), 353–363.

Misra, S. and H. S. Nair (2011). A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quantitative Marketing and Economics 9*(3), 211–257.

Nalbantian, H. R. and A. Schotter (1997). Productivity under group incentives: An experimental study. *The American Economic Review 87*(3), 314–341.

Reiss, P. C. and M. W. White (2008). What changes energy consumption? Prices and public pressures. *The RAND Journal of Economics 39*(3), 636–663.

Tan, J. H. and F. Bolle (2007). Team competition and the public goods game. *Economics letters 96*(1), 133–139.

Figure 1: Timeline of the Experiment

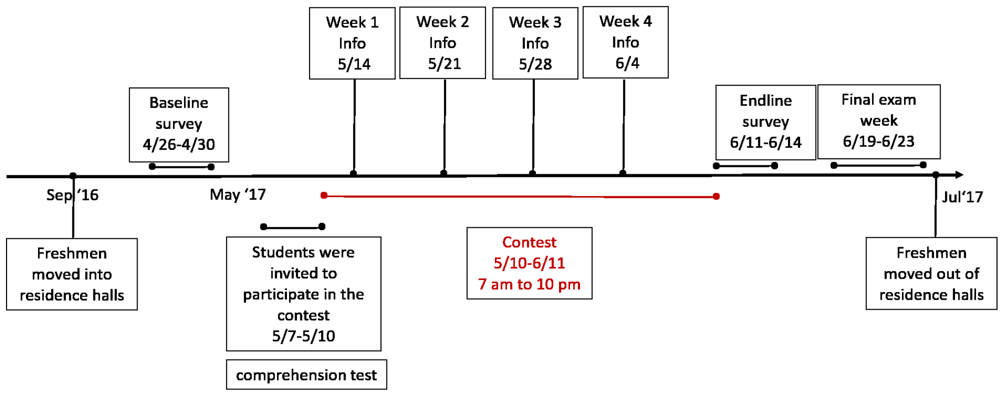


Figure 2: Usage distribution

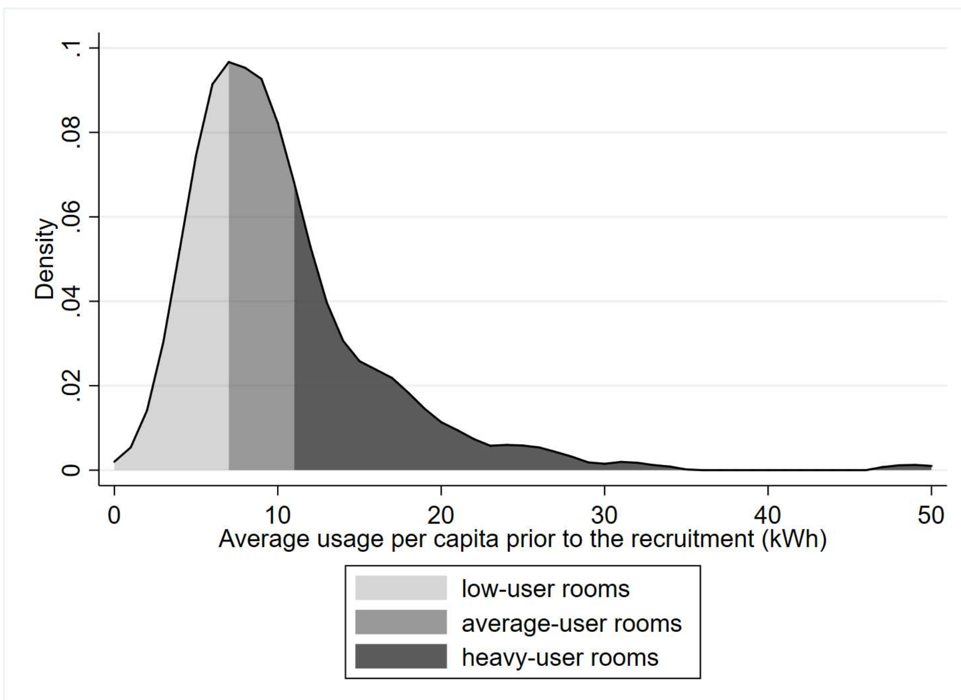
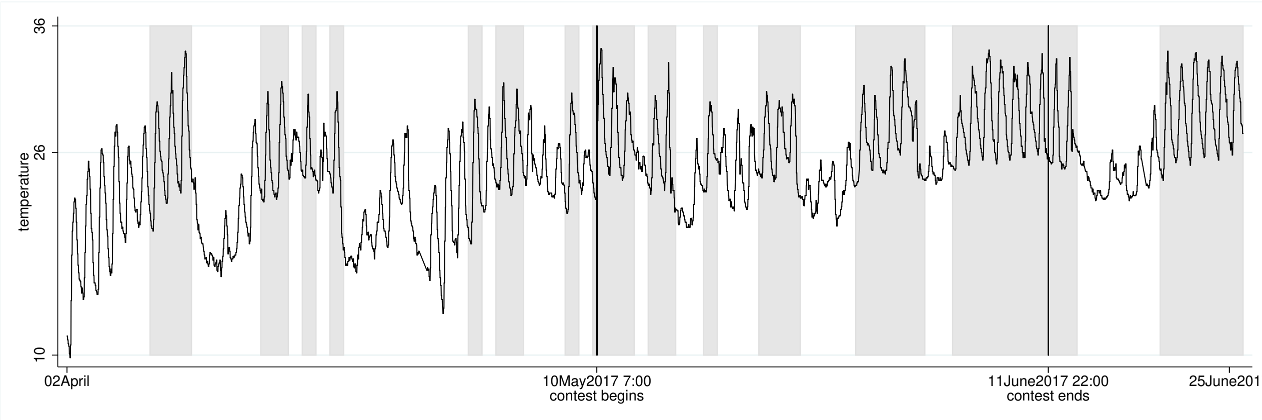
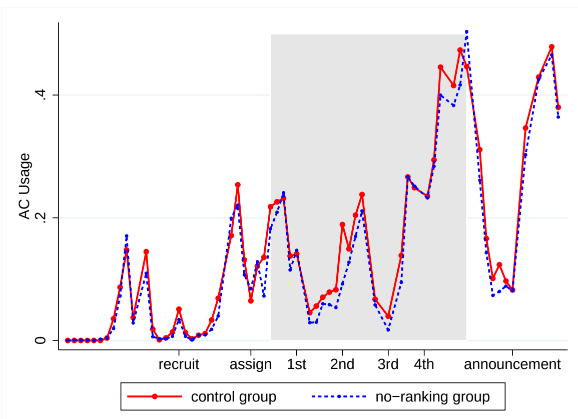


Figure 3: Hourly Temperatures

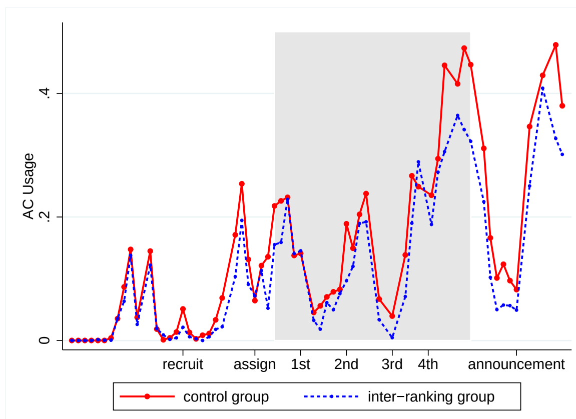


Note: The shaded area represents 'hot days' during the study period. These are days with an average temperature to be at least 28°C or a day with a maximum daily temperature to be at least 30°C.

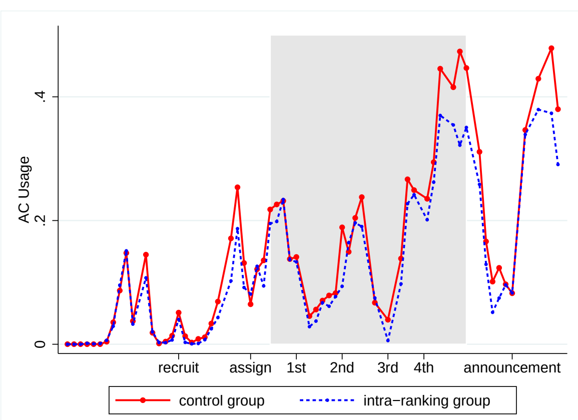
Figure 4: Average Daily AC Usage Per Room



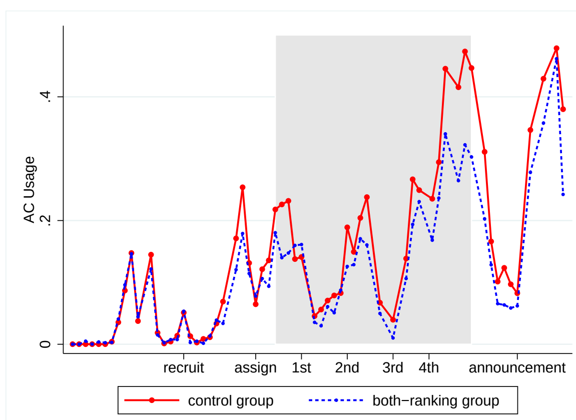
(a) No-ranking group



(b) Inter-ranking group



(c) Intra-ranking group



(d) Both-ranking group

Note: The shaded area represents the competition period.

Figure 5: Persistence of the Contest

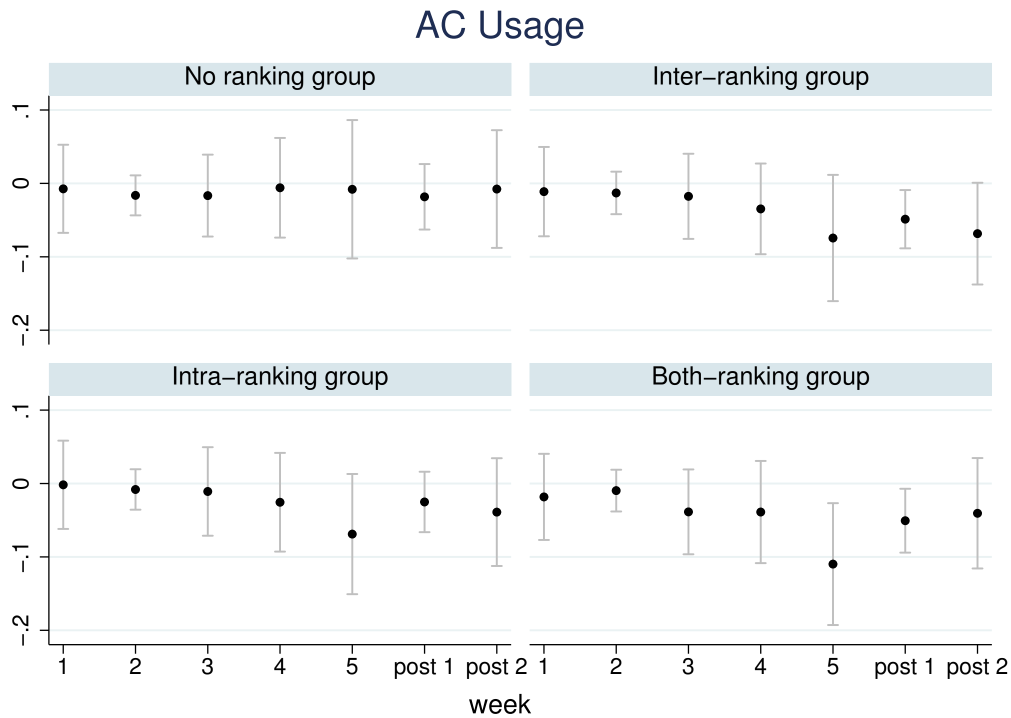


Table 1: Summary Statistics at the Room Level by Treatment Status

		Difference in means (compared to the control group)			
	(1) control mean (sd)	(2) no-ranking group (p-value)	(3) inter-ranking group (p-value)	(4) intra-ranking group (p-value)	(5) both-ranking group (p-value)
Panel A: Usage before the experiment					
Total usage	0.137	-0.023	-0.025	-0.013	-0.01
	(0.073)	(0.131)	(0.087)	(0.424)	(0.562)
Usage 220V	0.044	-0.006	-0.011	-0.006	-0.003
	(0.061)	(0.585)	(0.361)	(0.604)	(0.804)
Usage 110V	0.094	-0.016	-0.015	-0.007	-0.007
	(0.033)	(0.032)	(0.058)	(0.446)	(0.405)
Panel B: Usage during the contest					
Total usage	0.331	-0.037	-0.07	-0.046	-0.066
	(0.17)	(0.332)	(0.055)	(0.2)	(0.088)
Usage 220V	0.223	-0.021	-0.051	-0.04	-0.059
	(0.164)	(0.566)	(0.142)	(0.254)	(0.105)
Usage 110V	0.108	-0.017	-0.02	-0.007	-0.007
	(0.039)	(0.06)	(0.025)	(0.467)	(0.442)
Panel C: Room type					
Female	0.659	-0.081	-0.04	-0.088	-0.135
	(0.479)	(0.436)	(0.703)	(0.409)	(0.206)
4-bed rooms	0.818	-0.329	-0.271	-0.199	-0.247
	(0.390)	(0.001)	(0.006)	(0.04)	(0.012)

Notes: Total number of observations (rooms): 215. Control group: 44, no-ranking group: 45, inter-ranking group: 42, intra-ranking group: 42, both-ranking group: 42. All usage variables report hourly usage (kWh) per room.

Table 2: Beliefs about Intra-team Position

	Actual position		
	a	b	c
Share of self-report position			
No-ranking information			
$a^{'}$	0.63	0.30	0.08
$b^{'}$	0.23	0.32	0.63
$c^{'}$	0.14	0.38	0.29
Inter-ranking information			
$a^{'}$	0.61 (0.942)	0.42 (0.513)	0.23 (0.332)
$b^{'}$	0.24 (0.980)	0.27 (0.770)	0.47 (0.439)
$c^{'}$	0.15 (0.942)	0.31 (0.706)	0.30 (0.965)
Intra-ranking information			
$a^{'}$	0.90 (0.099)	0.06 (0.109)	0.03 (0.534)
$b^{'}$	0.10 (0.358)	0.76 (0.012)	0.17 (0.013)
$c^{'}$	0.00 (0.207)	0.18 (0.231)	0.81 (0.004)
Both-ranking information			
$a^{'}$	0.97 (0.018)	0.00 (0.021)	0.00 (0.328)
$b^{'}$	0.03 (0.097)	0.80 (0.005)	0.08 (0.002)
$c^{'}$	0.00 (0.169)	0.20 (0.261)	0.92 (0.000)

Notes: Number of respondents: 307. We construct indicator variables for each self-reported position, calculate their averages within a room, and report room averages in each cell. For cells corresponding to receiving ranking information, we calculate their differences in beliefs compared to cells receiving no ranking information, and report the p-values in parentheses. Positions refer to contestants' positions at the end of the contest. We use a, b, and c to denote a leading, middle, and behind position within a team, respectively. Similarly, we use $a^{'}$, $b^{'}$, and $c^{'}$ to denote self-reported within-team positions.

Table 3: Beliefs about Inter-team Position

	Actual position		
	A	B	C
Share of self-report position			
No-ranking information			
$A^{'}$	0.35	0.19	0.17
$B^{'}$	0.47	0.64	0.72
$C^{'}$	0.18	0.17	0.11
Inter-ranking information			
$A^{'}$	0.61 (0.197)	0.15 (0.764)	0.00 (0.153)
$B^{'}$	0.39 (0.707)	0.85 (0.221)	0.11 (0.001)
$C^{'}$	0.00 (0.038)	0.00 (0.161)	0.89 (0.000)
Intra-ranking information			
$A^{'}$	0.23 (0.541)	0.33 (0.373)	0.15 (0.917)
$B^{'}$	0.52 (0.812)	0.46 (0.289)	0.48 (0.155)
$C^{'}$	0.25 (0.702)	0.21 (0.762)	0.37 (0.083)
Both-ranking information			
$A^{'}$	0.79 (0.015)	0.19 (1.000)	0.00 (0.063)
$B^{'}$	0.21 (0.154)	0.64 (1.000)	0.20 (0.001)
$C^{'}$	0.00 (0.052)	0.17 (1.000)	0.80 (0.000)

Notes: Number of respondents: 307. We construct indicator variables for each self-reported position, calculate their averages within a room, and report room averages in each cell. For cells corresponding to receiving ranking information, we calculate their differences in beliefs compared to cells receiving no ranking information, and report the p-values in parentheses. Positions refer to contestants' positions at the end of the contest. We use A, B, and C to denote a leading, middle, and behind position across teams, respectively. Similarly, we use $A^{'}$, $B^{'}$, and $C^{'}$ to denote self-reported inter-team positions.

Table 4: Impact of Contest and Information Treatment on Usage

	All days		Hot days	
	(1) Total Usage	(2) AC usage	(3) Total Usage	(4) AC usage
β_0: 1(Week 1)×1(In Contest)	-0.005	-0.009	-0.003	-0.013
	(0.026)	(0.025)	(0.026)	(0.026)
β_1: 1(Info weeks)×1(No ranking)	-0.013	-0.013	0.001	-0.004
	(0.027)	(0.027)	(0.033)	(0.033)
β_2: 1(Info weeks)×1(Inter ranking)	-0.047^{+}	-0.040	-0.052	-0.050
	(0.027)	(0.027)	(0.033)	(0.031)
β_3: 1(Info weeks)×1(Intra ranking)	-0.037	-0.036	-0.046	-0.051
	(0.026)	(0.027)	(0.032)	(0.032)
β_4: 1(Info weeks)×1(Both rankings)	-0.058*	-0.057*	-0.081*	-0.083*
	(0.027)	(0.028)	(0.031)	(0.031)
Room-by-day-of-week fixed effects	Yes	Yes	Yes	Yes
Observations	141255	141255	65790	65790

Notes: The unit of hourly electricity usage is kilowatt hours (kWh). 'Hot days' are days with an average temperature of at least 28°C or a maximum daily temperature of at least 30°C. All standard errors are clustered at the team level. $^{+}$ $p < 0.10$, * $p < 0.05$

Table 5: Survey Evidence: Did Weekly Text Messages Affect Your Energy Use?

	(1) β_1: 1(No ranking)	(2) β_2: 1(Inter ranking)	(3) β_3: 1(Intra ranking)	(4) β_4: 1(Both rankings)	(5) Constant
Coefficients	0.020	0.178*	0.143^{+}	0.240**	0.178**
Standard errors	(0.062)	(0.073)	(0.076)	(0.072)	(0.033)

Notes: Total observation is 399. The dependent variable is an indicator variable that equals one when a participant reports that information is effective in changing her/his energy use and zero otherwise. All standard errors are clustered at the team level. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 6: Alternative Specifications

	(1) AC Usage	(2) AC usage	(3) AC Usage	(4) AC usage	(5) AC usage
β_0: 1(Week 1)×1(In Contest)	-0.012	-0.014	-0.012	0.006	-0.010
	(0.022)	(0.022)	(0.026)	(0.031)	(0.023)
β_1: 1(Info weeks)×1(No ranking)	-0.012	-0.013	-0.002	0.046	-0.006
	(0.027)	(0.027)	(0.033)	(0.043)	(0.026)
β_2: 1(Info weeks)×1(Inter ranking)	-0.043^{+}	-0.043^{+}	-0.051	-0.041	-0.045
	(0.025)	(0.025)	(0.032)	(0.039)	(0.027)
β_3: 1(Info weeks)×1(Intra ranking)	-0.041	-0.042	-0.054	-0.040	-0.044
	(0.027)	(0.027)	(0.033)	(0.042)	(0.027)
β_4: 1(Info weeks)×1(Both rankings)	-0.066*	-0.067*	-0.085**	-0.092*	-0.071*
	(0.028)	(0.027)	(0.032)	(0.042)	(0.027)
Fixed effects	room	room-by-hour-of-day	room-by-day-of-week-by-time-of-day	room-by-day-of-week	room-by-day-of-week
4-bed room only?	No	No	No	Yes	No
arcsinh transformation?	No	No	No	No	Yes
Hot days?	Yes	Yes	Yes	Yes	Yes
Observations	65790	65790	65790	40086	65790

Notes: The unit of hourly electricity usage is kilowatt hours (kWh). Time-of-day fixed effects are fixed effects for morning (7:00-11:59), afternoon (12:00-17:59), and night (18:00-21:59). The inverse hyperbolic sine (arcsinh) transformation of a variable x is defined as $\text{arcsinh}(x) = \ln(x + \sqrt{x^2+1})$. 'Hot days' are days with an average temperature of at least 28°C or a maximum daily temperature of at least 30°C. All standard errors are clustered at the team level. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 7: Electricity Saved in the Contest

	(1) Hourly Usage	(2) Daily Usage (Interpolation)
Compete in the contest	-0.032	-0.425
p-value	(0.153)	(0.113)
Observations	141255	12470
Estimated counterfacutal usage (no contest)	0.304	4.197
Estimated saving	10.6%	10.1%
# of rooms in contest	171	171
# of days (hours)	311	31
Estimated saving (kWh)	1712	2251

Appendices

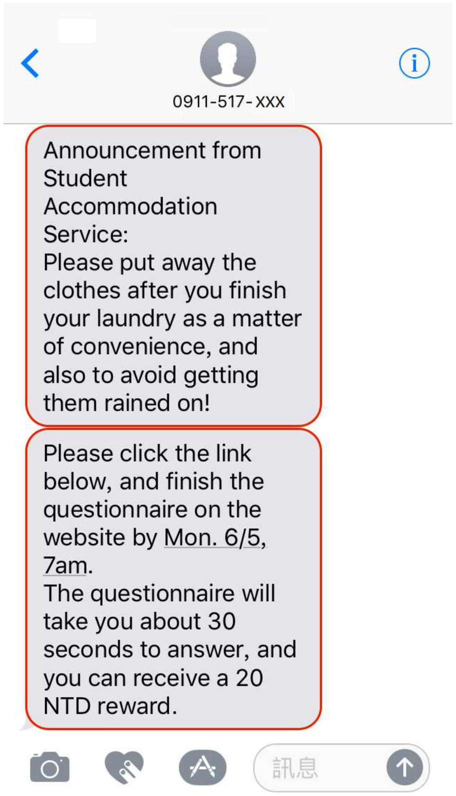
A Materials from the Field Experiment

In Table A1, we provide the experimental design and the number of recruited participants in each treatment group. Figure A1 shows sample text messages sent to participants with different treatment assignments.

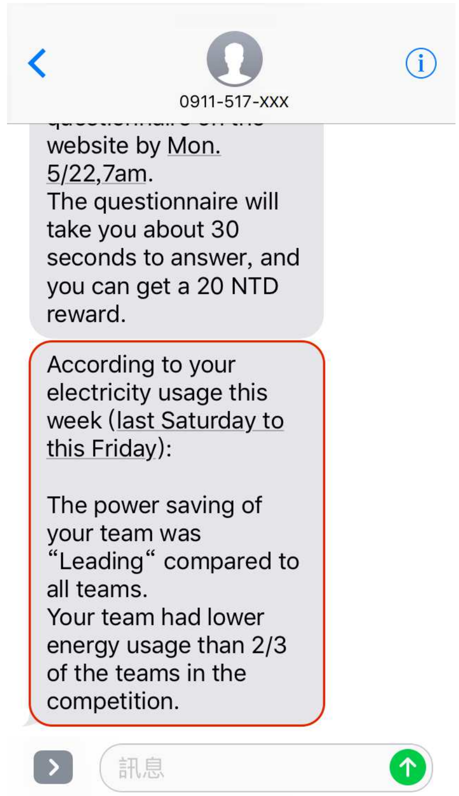
We calculated the power of our experiment using the AC usage from May 10, 2016 to June 11, 2016 (one year before the contest). We also restricted the sampling hours to be between 7:00 a.m. and 10:00 p.m. (same as the competition window of the contest) and in 'hot days'.[29] The mean and the standard deviation of the AC usage are 0.237 kWh and 0.398 kWh, respectively. Because we observed multiple decisions (one for each hour) from each room, we calculated the interclass correlation coefficient (ICC, denoted as ρ) to reflect correlations within the room. The resulting ICC is 0.082. The power (1-β) for a two-sample means test under a 5% significance level ($\alpha = 0.05$) with 45 rooms in the control group and 45 rooms in the treatment group for 265 hours (number of hours in the 'contest window' during hot days in 2016) is 0.904 and 0.815 when the effect size of the treatment is -0.08 kWh, and -0.07 kWh respectively. Unfortunately, during the actual contest, we only had 228 hours with recorded usage during hot days, we present the power calculations under 228 hours in Figure A2.

[29]Recall that hot days were defined as having an average temperature of at least 28°C or a maximum daily temperature of at least 30°C.

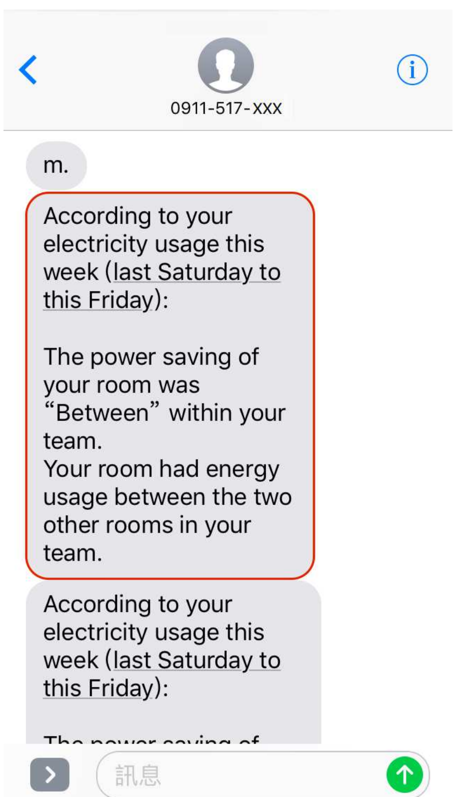
Figure A1: Text Messages for Control and Treatment Groups (Translated to English)



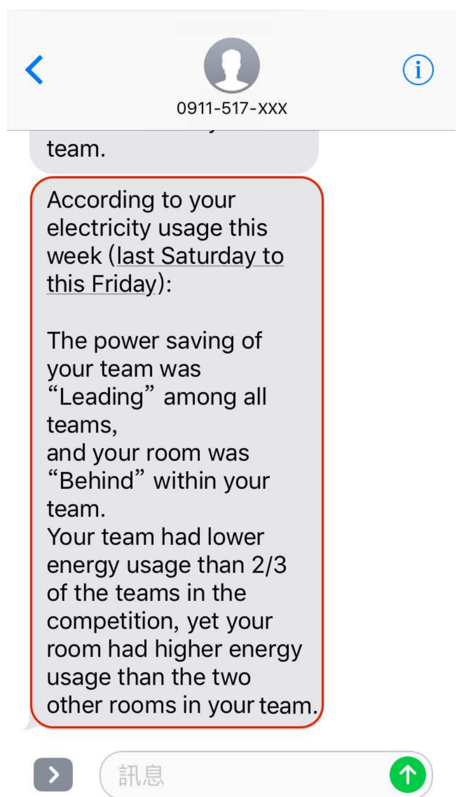
(a) Control and no-ranking group



(b) Inter-ranking group



(c) Intra-ranking group

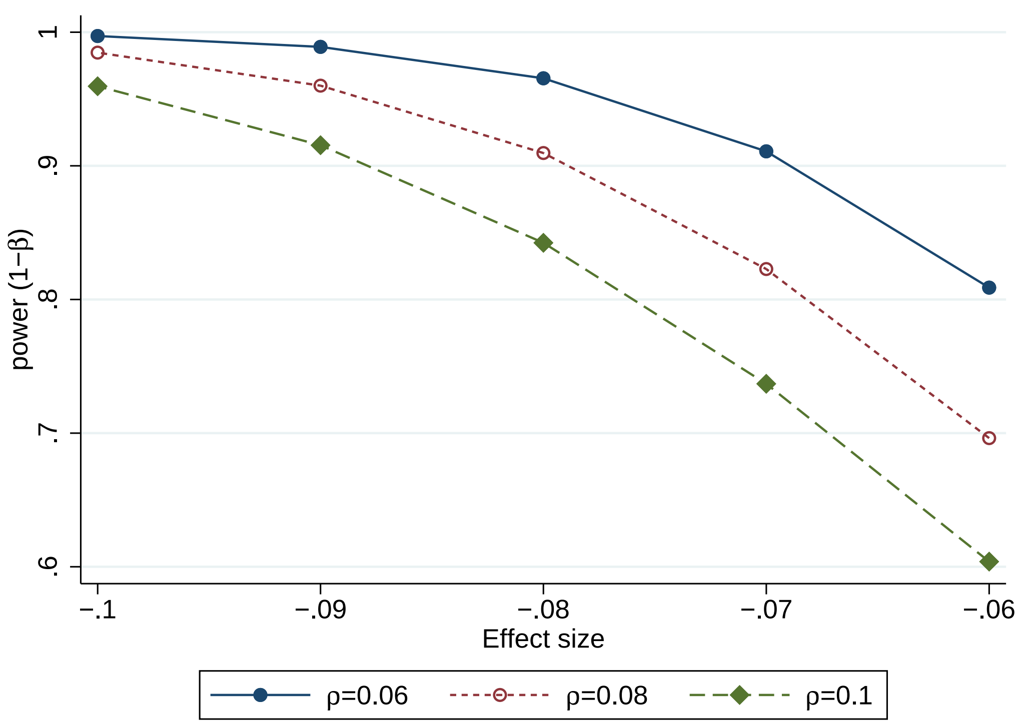


(d) Both-ranking group

Table A1: Experimental Design

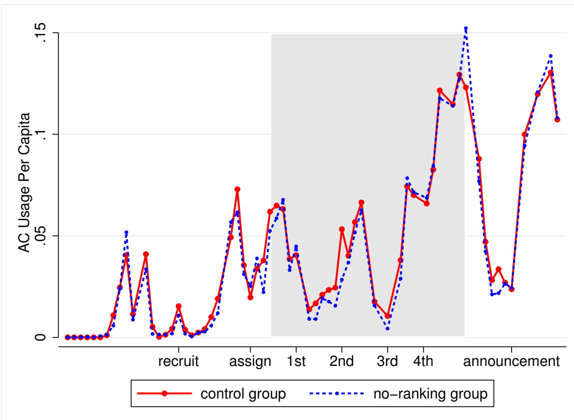
	Number of participants	In contest?	Information about their team's relative performance across teams?	Information about their room's relative performance within the team?
Control group	118	No	No	No
No-ranking treatment	111	Yes	No	No
Inter-ranking treatment	99	Yes	Yes	No
Intra-ranking treatment	117	Yes	No	Yes
Both-ranking treatment	108	Yes	Yes	Yes

Figure A2: Effect Size and Power of the Experiment

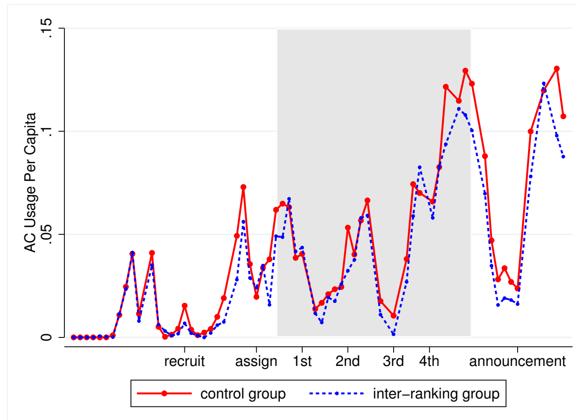


B Additional Results and Robustness Checks

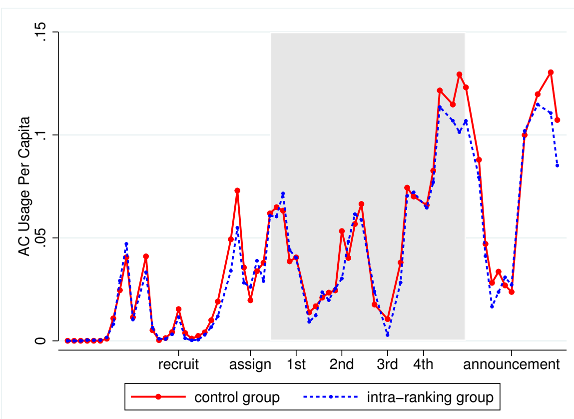
Figure B1: Average Daily AC Usage Per Capita



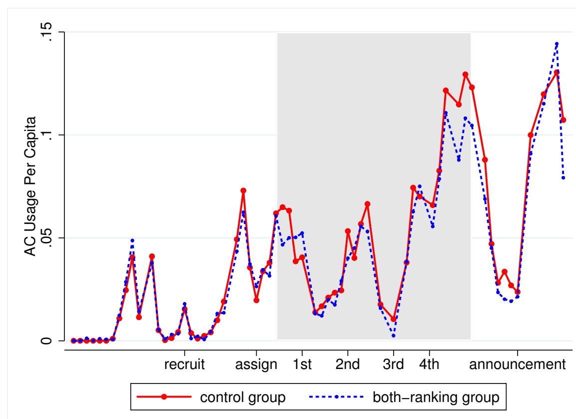
(a) No-ranking group



(b) Inter-ranking group



(c) Intra-ranking group



(d) Both-ranking group

Note: The shaded area represents the competition period.

Table B1: Summary Statistics of Participants from the Baseline Survey

	(1) control mean (sd)	Difference in means (compared to the control group) (2) no-ranking group (p-value)	(3) inter-ranking group (p-value)	(4) intra-ranking group (p-value)	(5) both-ranking group (p-value)
Effort	5.212	0.022	-0.293	-0.297	0.242
	(2.305)	(0.939)	(0.317)	(0.31)	(0.427)
Credits	19.771	0.328	0.734	0.6	0.655
	(3.793)	(0.465)	(0.12)	(0.154)	(0.148)
Part-time hours	3.309	-0.426	0.817	0.768	0.107
	(8.213)	(0.639)	(0.43)	(0.596)	(0.91)
Any club	0.466	0.011	-0.001	-0.022	-0.003
	(0.501)	(0.864)	(0.983)	(0.74)	(0.963)
Days in dorm	5.915	-0.285	-0.097	-0.061	-0.124
	(1.051)	(0.067)	(0.539)	(0.669)	(0.405)
Bill	1215	-24	-37	-25	-35
	(450)	(0.662)	(0.581)	(0.706)	(0.571)
Pay by oneself	0.068	0.022	0.064	-0.017	-0.022
	(0.252)	(0.533)	(0.116)	(0.595)	(0.49)
Pay by family	0.89	-0.016	-0.041	0.016	0.008
	(0.314)	(0.71)	(0.368)	(0.684)	(0.84)
Pay by loan	0.042	0.003	-0.022	-0.008	0.013
	(0.202)	(0.922)	(0.36)	(0.745)	(0.647)
Satisfactory level	7.398	-0.362	-0.57	0.209	0.167
	(1.690)	(0.1)	(0.022)	(0.303)	(0.447)

Notes: Total number of participants: 553. Effort (Provided you and your roommates need 10 units of effort to finish the room chores, from 1 to 10, which number best describes the effort you put into the work?); Credits (How many credit hours are you taking this term?); Part-time hours (How many hours do you spend on part-time jobs every week on average?); Any club? (Please provide the name of any club or student association which you have participated in. If you did not participate in any club or student association, please write ‘None’.); Days in dorm (How many days per week do you stay in the room?); Bill (How much did you pay for the dorm electricity bill?); Pay by oneself/Pay by family/Pay by loan (How did you pay for your dormitory fee?) Satisfactory level (Please give a score for your dorm according to your life in the dorm so far. 1: the worst to 10: the best.)

Table B2: Results without Using Pre-Experiment Data

	All days		Hot days	
	Total	AC	Total	AC
β_0: 1(Week 1)×1(In Contest)	-0.025	-0.019	-0.028	-0.022
	(0.029)	(0.029)	(0.030)	(0.030)
β_1: 1(Info weeks)×1(No ranking)	-0.037	-0.021	-0.034	-0.018
	(0.033)	(0.033)	(0.040)	(0.040)
β_2: 1(Info weeks)×1(Inter ranking)	-0.070*	-0.051	-0.080*	-0.060
	(0.031)	(0.033)	(0.038)	(0.039)
β_3: 1(Info weeks)×1(Intra ranking)	-0.046	-0.040	-0.055	-0.048
	(0.033)	(0.034)	(0.041)	(0.041)
β_4: 1(Info weeks)×1(Both rankings)	-0.066^{+}	-0.059^{+}	-0.084*	-0.076^{+}
	(0.034)	(0.036)	(0.041)	(0.042)
Observations	66865	66865	49020	49020

Notes: This table reports estimation results without using pre-experiment data. This specification does not allow for room-level fixed effects and cannot remove pre-experiment permanent differences in usage. The unit of hourly electricity usage is kilowatt hours (kWh). 'Hot days' are days with an average temperature of at least 28°C or a maximum daily temperature of at least 30°C. All standard errors are clustered at the team level. $^{+}$ $p < 0.10$, * $p < 0.05$

Table B3: Results without Using Pre-Experiment Data (With Random Effects)

	All days		Hot days	
	Total	AC	Total	AC
β_0: 1(Week 1)×1(In Contest)	-0.025	-0.019	-0.028	-0.022
	(0.029)	(0.029)	(0.030)	(0.030)
β_1: 1(Info weeks)×1(No ranking)	-0.036	-0.026	-0.039	-0.029
	(0.029)	(0.029)	(0.034)	(0.035)
β_2: 1(Info weeks)×1(Inter ranking)	-0.061*	-0.045	-0.069^{+}	-0.052
	(0.031)	(0.031)	(0.035)	(0.036)
β_3: 1(Info weeks)×1(Intra ranking)	-0.055^{+}	-0.046	-0.064^{+}	-0.056
	(0.029)	(0.030)	(0.034)	(0.035)
β_4: 1(Info weeks)×1(Both rankings)	-0.068*	-0.054^{+}	-0.080*	-0.065^{+}
	(0.029)	(0.030)	(0.035)	(0.036)
Observations	66865	66865	49020	49020

Notes: This table reports estimation results without using pre-experiment data. This specification does not allow for room-level fixed effects and cannot remove pre-experiment permanent differences in usage. We estimate the coefficients using a random effects model instead. The unit of hourly electricity usage is kilowatt hours (kWh). 'Hot days' are days with an average temperature of at least 28°C or a maximum daily temperature of at least 30°C. All standard errors are clustered at the team level. $^{+}$ $p < 0.10$, * $p < 0.05$

C Heterogeneity in the Effects of Competitive Information, Social Information, and Their Interactions

In this subsection, we examine whether the effects of providing ranking information are heterogeneous across inter- and intra-team positions. To this end, we estimate the model by splitting samples according to room states at the end of the first week and test the differences between estimated coefficients. While splitting samples allows for examining a rich set of heterogeneous treatment effects, we suffer from using a smaller sample size and losing statistical power in many cases. Nevertheless, in most cases, once we incorporate the panel nature of our data set, the power of the test statistics (the difference between coefficients) exceeds 0.8 when the test statistics are greater than 0.03.[30]

The results are presented in Table C1. Each column in Table C1 provides estimates of equation (1) by splitting the sample based on relative room position at the end of the first contest week (when all participants had not yet received any information). Because our information treatments kicked in at the end of the first week, we expect to find behavioral responses after the first week if treatments were effective. Columns (1)–(3) in Table C1 split the sample by inter-team positions (A, B, C), whereas columns (4)–(6) split the sample by intra-team positions (a, b, c). By comparing estimated coefficients conditional on relative room position, we can test whether the treatment effects varied by relative position. Throughout the columns, the base group used includes all participants from the control group.

First, we find that electricity savings for position A were generally the largest, followed by those for position B. Those for position C were the smallest. These results suggest that teams leading in the first week continued to lead throughout the contest. For leading teams

[30] We assume that the true data-generating process follows equation (1) and conduct the power analysis by simulation. In each simulation, we use estimates from Table C1 to create samples of the same size used in Table C1, each consisting of a subset of rooms (by position at the end of the first week) and 306 hours during the study period. We adjust the difference between parameters and find the power for all tests ($\beta_2 - \beta_1 = 0, \beta_3 - \beta_1 = 0, \beta_4 - \beta_1 = 0, \beta_4 - \beta_2 = 0, \beta_4 - \beta_3 = 0$).

Table C1: Impact of Different Treatments on Usage by Position

	AC usage on hot days by position at the end of the first week					
	(1)	(2)	(3)	(4)	(5)	(6)
	A	B	C	a	b	c
β_0: 1(Week 1)×1(In Contest)	-0.060*	-0.028	0.049	-0.077**	-0.001	0.039
	(0.027)	(0.030)	(0.032)	(0.024)	(0.030)	(0.033)
β_1: 1(Info weeks)×1(No ranking)	-0.058^{+}	-0.004	0.076	-0.143**	0.026	0.106*
	(0.029)	(0.061)	(0.050)	(0.033)	(0.043)	(0.043)
β_2: 1(Info weeks)×1(Inter ranking)	-0.123**	-0.013	-0.021	-0.150**	-0.012	0.014
	(0.032)	(0.032)	(0.063)	(0.032)	(0.044)	(0.050)
β_3: 1(Info weeks)×1(Intra ranking)	-0.075	-0.106**	0.036	-0.123**	-0.025	-0.005
	(0.050)	(0.025)	(0.048)	(0.036)	(0.047)	(0.034)
β_4: 1(Info weeks)×1(Both rankings)	-0.138**	-0.131*	-0.018	-0.154**	-0.068^{+}	-0.026
	(0.031)	(0.050)	(0.038)	(0.033)	(0.038)	(0.041)
Room-by-day-of-week fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30906	30906	30906	30906	30906	30906

Notes: The unit of hourly electricity usage is kilowatt hours (kWh). Positions refer to contestants' positions at the end of the first week. A: leading across teams; B: in the middle across teams; C: behind across teams; a: leading within the team; b: in the middle within the team; c: behind within the team. 'Hot days' are days with an average temperature of at least 28°C or a maximum daily temperature of at least 30°C. All standard errors are clustered at the team level. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

(position A), we can reject the hypothesis that inter-ranking information was irrelevant in the contest ($\beta_2 - \beta_1 = 0$) at the 5% level (with a p-value of 0.029), but we cannot do so for teams in position B or C. The results suggest that once inter-team ranking information was provided after the first week, those who were told that they were in the winning position started to exert more effort than their counterparts, who were in a similar winning position but were not given such information.

Turning to columns associated with intra-team positions (a, b, c), we find that the savings in electricity from low users at the end of the first week (position a) were the largest and those from heavy users (position c) were the smallest, indicating that low users tended to keep using less electricity than heavy users after the first week. To test whether social information promoted additional effort in the contest, we calculate ($\beta_3 - \beta_1$) as our test statistic, and a negative and significant result indicates that social information helped contestants save more electricity. Examining column (4) (position a), we find that compared to average users in the control group, contestants without ranking information reduced their average hourly AC usage by 0.143 kWh; contestants receiving intra-ranking information only reduced their hourly AC usage by 0.123 kWh. We thus cannot reject the hypothesis that providing social information for team leaders in a group contest helped them exert more effort in our setting.[31] For rooms in position b, the result also does not support the effectiveness of social information: the test statistic (p-value) associated with the hypothesis is -0.051 (0.330). Finally, for rooms in position c, the test statistic (p-value) for the hypothesis that social information is irrelevant is -0.111 (0.006). Therefore, our results find that providing intra-team ranking information to team laggards successfully helped them exert more effort.

The above results are consistent with the literature findings that social information is more effective for changing behaviors among high electricity users (Allcott, 2011; Delmas and Lessem, 2014; Ayres et al., 2013; Kandul et al., 2020). For low users, our estimated coefficient suggests that they seemed to increase their usage after receiving social information,

[31] The test statistic (p-value) for the hypothesis ($\beta_3 - \beta_1 = 0$) is 0.020 (0.582).

but the effect was not precisely estimated. Notably, unlike heavy users, who tended to be over-confident about their intra-team positions without information, low users were more likely to know that they consumed less electricity than average users; thus, providing social information for low users was less likely to correct their biased beliefs about intra-team positions.

Next, we explore the interaction effects of receiving inter- and intra-team ranking information. We do so by comparing coefficients from the both-ranking group to those from the inter- or intra-ranking group. First, we ask whether it was helpful to provide additional social information for those who already knew their competitive states (A, B, C). The test statistic for this question is $(\beta_4 - \beta_2)$, using the coefficients from columns (1)–(3) in Table C1. The associated p-values are 0.605, 0.024, and 0.968 for positions A, B, and C, respectively. Thus, we find evidence to support that providing social information for teams in the middle position helped them exert more effort. Next, we ask whether it was helpful to provide additional competitive information for those who already knew their social states (a, b, c). The test statistic for this question is $(\beta_4 - \beta_3)$, using the coefficients from columns (4)–(6) in Table C1. All of the test statistics are negative, but the associated p-values are 0.404, 0.362, and 0.598, for positions a, b, and c, respectively; therefore, we do not have enough evidence to support that additional competitive information was effective for contestants once they obtained social information about their relative performance within a group. Nevertheless, we do not find significant dampening effects on efforts to provide additional social information to those who knew their competitive states or efforts to provide competitive information to those who knew their social states.