Timing is Everything: Adding Peak-Focused Feedback to Monthly Home Energy Reports

Dan Fredman, Oracle Utilities
Delaney Tight, Oracle Utilities

ABSTRACT

Home Energy Reports (HERs) produce consistent energy savings via bill-period normative comparisons (NCs), but time-of-use (TOU) rate deployments demand behavioral tools that target *when* customers consume, not just how much. This paper evaluates Peak-Focused (PF) HERs, a product feature replacing billing-period comparisons with on-peak NCs and related peak insights, including a "close the gap" (CTG) insight, quantifying the annualized cost of a customer's peak usage relative to neighbors. In Spring 2025, a municipal utility in California deployed a randomized controlled trial (RCT) across 313,417 residential TOU customers, including a legacy cohort with prior HER exposure and a new cohort with none. Nine months of advanced meter interval data and billing usage were analyzed using lagged-dependent variable regression. PF HERs produced statistically significant on-peak reductions in the legacy cohort (up to 0.52%) when compared to TOU HERs, with effects accumulating gradually rather than spiking at first exposure — consistent with habit formation over repeated exposure. Program-naive customers receiving either TOU or PF HER saw average peak savings of 0.7% relative to control groups, reflecting strong first-year responses. TOU HER drove a higher maximum peak reduction of 1.51%, while PF HER saved at 1.36%. The CTG cost-framing insight produced no statistically significant incremental peak savings, partly due to under-triggering from business logic thresholds. No arm produced significant differences in total monthly energy use, suggesting PF NC and bill-based NC have similar monthly savings impacts. These findings offer replicable design guidance for utilities pursuing load flexibility through existing behavioral infrastructure.

Introduction

Traditional demand-side management (DSM) programs have delivered substantial energy savings through equipment rebates and efficiency upgrades, but these programs address aggregate consumption rather than its temporal distribution. As utilities deploy time-varying rates (TVRs) to align customer consumption with system costs and renewable generation patterns, they need complementary behavioral interventions that help customers respond to these price signals. Home Energy Reports (HERs)—the well-established behavioral energy efficiency (BEE) program delivering social normative comparisons via mail and email—have produced consistent energy savings of 1–2% across millions of households (Allcott 2011; Allcott and Rogers 2014). While a high proportion of HER savings have been found to often occur during peak hours (Guidehouse 2025), these programs are not designed with the sole focus of reducing energy during specific time periods, and most behavioral energy efficiency research has examined interventions designed to reduce total consumption, not shift it across time (Vine, Buys, and Morris 2013). This mismatch between the temporal precision of modern rate structures and the aggregate focus of behavioral programs represents both a challenge and an opportunity for utilities pursuing load flexibility alongside efficiency goals as they adopt Advanced Metering Infrastructure (AMI) and time-of-use (TOU) rates. In fact, research has found without clear feedback about peak-period consumption patterns and their financial implications, customers may not respond to temporal price signals even when rates are well-designed because they lack the best information to respond effectively (Nicolson, Fell, and Huebner 2018; Wolak 2011).

This paper examines whether behavioral interventions can be tailored to target peak-period consumption through Peak-Focused HER (PF HER), a new product feature built within the standard HER platform. We test whether providing customers with peer comparisons focused specifically on their on-peak energy use (rather than total billing-period consumption) drives meaningful peak load reduction, and whether augmenting those comparisons with a loss-framed financial insight that quantifies cost savings from matching neighbors' peak-period behavior further amplifies that effect. The intervention runs through the same distribution channels and analytical infrastructure already in place for standard HER programs, making it a low-cost enhancement to existing behavioral portfolios.

PF HER occupies a distinct position among behavioral DSM products. Behavioral Demand Response (BDR) programs deliver event-based messaging during specific critical peak periods, demonstrating meaningful reductions during targeted events but providing minimal ongoing education about typical peak-period patterns. Behavioral Load Shaping (BLS) programs use recurring messages to encourage consumption shifts, but without a savings-oriented message (i.e., the emphasis is shift-and-save rather than conserve-and-save). PF HER combines the persistent, recurring structure of standard HER with the temporal specificity that TOU rate structures require, filling a gap rather than replacing existing behavioral tools.

This research addresses four questions: Can normative comparisons emphasizing peak-period consumption drive measurable on-peak reductions? Does adding financial cost framing (what we later refer to as 'close the gap' or CTG messaging, the PF$ treatment arm) amplify those effects beyond peak-focused feedback alone? Do peak-focused interventions maintain total energy savings, or does the emphasis on peak hours merely shift consumption without reducing it? Given answers to these questions, what are the practical implications for utilities seeking to balance behavioral load flexibility with efficiency objectives under expanding AMI and TVR deployments?

We present four main contributions. First, this is the first large-scale field evaluation of normative comparisons targeting peak-period consumption rather than billing-period totals. Second, we provide novel examination of the interaction between social norms and financial cost framing in the context of load flexibility with implications for future use. Third, we show that a single behavioral intervention can pursue dual objectives — peak load reduction and energy conservation — through existing HER infrastructure.

Literature Review and Theoretical Framework

Connecting energy efficiency to load flexibility. DSM implementation has undergone a significant reorientation over the past decade. Grid operators and utilities now face a distinct challenge: not just how much customers consume, but when. Rapid growth in variable renewable generation has made the temporal shape of residential demand a first-order operational concern. A 2021 Lawrence Berkeley National Laboratory inventory found 148 demand flexibility programs and 93 time-varying rate structures operating across the United States, underscoring both the breadth of utility commitment to this shift and the diversity of program designs being tested (Murphy, Miller, and Deason 2021).

Social Norms and Energy Feedback. HERs represent the most extensively evaluated behavioral intervention in residential energy efficiency. Allcott (2011) established the baseline across roughly 600,000 households: billing-based social comparisons produced average energy savings of about 2 percent, concentrated in high-consuming households. A meta-analysis of post-2018 Opower program evaluations (Munson 2022) found average electric savings of 1.16% and a positive correlation between years of report receipt and savings magnitude, consistent with the accumulating effect documented in Allcott and Rogers (2014).

Cialdini, Reno, and Kallgren's (1990) focus theory of normative conduct holds that the behavioral relevance of a norm depends on how salient and proximate it is to the decision context. This has a direct implication for load flexibility. Billing-period normative comparisons aggregate consumption across all hours of the day and all days of a billing cycle. They make total usage socially visible but leave peak-period behavior entirely invisible. A customer who has low monthly usage, but heavy peak usage, receives a favorable comparison, which provides no signal to shift load. Vine, Buys, and Morris (2013) explicitly identified peak-period behavior as underexplored territory in the feedback literature. Our internal investigation leading up to the creation of this new HER variant confirmed that assessment—customers often live in a 'confusion matrix' where their on-peak usage may be worse while their bill period usage is better than their normative cohort. Fischer (2008) and Karlin, Zinger, and Ford (2015) each established that feedback effectiveness depends substantially on specificity and timing. The implication is direct: a normative comparison designed to make peak-period behavior socially visible may activate the same social influence mechanisms that billing-based HERs rely on but direct them toward the behavioral change utilities now need most.

Financial Framing and Cost Salience. Normative comparisons create awareness of relative standing but do not, by themselves, communicate the financial stakes of peak-period consumption under time-varying rates. Customers on TOU rates face a situation where the same kilowatt-hour (kWh) can cost two to five times more depending on when it is consumed, yet most customers do not carry an accurate mental model of how their peak-period behavior maps onto their bill. Attari, DeKay, Davidson, and Bruine de Bruin (2010) found that consumers underestimate the energy consumption of high-intensity appliances by a factor of 2.8 on average, with the largest errors concentrated in high-energy activities that dominate peak periods. Fredman, Palchak, Koliba, and Zia (2018) found that combining real-time feedback with financial incentives in rental households corrected participants' misperceptions of large appliance energy consumption and doubled savings relative to feedback alone, suggesting that perception correction and financial salience interact to drive behavior change.

Two behavioral mechanisms help explain why quantifying peak-period costs in a normative comparison should amplify its effectiveness. First, prospect theory (Kahneman and Tversky 1979) established that people weight losses more heavily than equivalent gains, typically at a ratio of 1.5 to 2.5. Framing the gap between a customer's peak-period costs and those of efficient neighbors as money being lost activates loss aversion in a way that neutral comparative feedback does not. Kesternich, Lange, and Sturm (2019) tested this directly in a field experiment on electricity saving behavior and found that loss-framed incentives produced approximately 5 percent savings, outperforming gain-framed conditions. Second, mental accounting theory (Thaler 1999) predicts that people manage expenditures through categorical mental budgets. Hahnel, Ortmann, Korcaj, Spada, and Brosch (2020) demonstrated that households maintain separate mental accounts for different categories of household expenses, with energy costs typically treated as a largely uncontrollable fixed outlay. Making peak-period charges explicit and socially comparative may activate a distinct mental account for peak-period consumption that does not exist when customers only see a billing-period total.

Theoretical Framework. PF HER operates through three behavioral mechanisms that billing-based HERs do not engage:

1. Temporal norm salience: constructing a normative comparison around on-peak consumption makes peak-period behavior visible as a distinct social category, consistent with Cialdini's focus theory prediction that norm salience is a precondition for normative influence (Cialdini, Reno, and Kallgren 1990).
2. Perception correction: customers receiving feedback about their relative standing on peak usage and precise guidance on what actions will influence that usage gain accurate information about a consumption domain where misperceptions are systematic and consequential (Attari et al. 2010).
3. Loss-framed cost salience (present in the PF$ treatment group of our experiment): quantifying the peak-period cost differential between the customer and efficient neighbors in dollar terms activates loss aversion and may open a peak-specific mental account that makes the cost of peak consumption tractable and personally relevant.

From these mechanisms and context, we derive four testable hypotheses. H1: Customers receiving peak-focused normative comparisons will exhibit greater reductions in on-peak consumption than customers receiving standard billing-based HER, as peak-specific norms direct attention and social influence toward precisely the behavior time-varying rates are designed to shift. H2: Customers receiving the CTG financial insight in addition to the peak-focused comparison will exhibit larger peak reductions than customers receiving the peak comparison alone, as loss-framed cost framing amplifies norm salience through financial stakes. H3: Peak-focused interventions will maintain or increase total energy savings relative to standard HER, because the behavioral mechanisms driving peak reduction are not peak-period-specific in their psychological effects and may generalize to total consumption. H4: In utilities with a mature HER environment (legacy recipients), new recipients with no prior HER exposure will respond more strongly to the peak-focused treatment, consistent with the habituation effects documented in Allcott and Rogers (2014).

Methods

Features of Standard, Time-of-Use, and Peak-Focused HERs. HERs deliver personalized energy feedback through a set of print and email modules (Oracle n.d.). The front page centers on a normative comparison (NC) showing the customer's billing-period consumption against a matched peer group with an adjacent injunctive norm (Fair, Good, or Great) indicating where they fall in the distribution. The back page pairs a self-comparison insight with a tips module that surfaces recommended actions using customer-specific business logic. The TOU HER preserves the billing-period NC and injunctive norm on the front page but replaces the back-page self-comparison and standard tips with rate education and peak-conscious tips. A TOU 101 module provides conceptual grounding in peak and off-peak pricing, and the tips module rotates formats designed to shift appliance use away from peak hours. Motivation and rate education run in parallel: the NC targets total kWh, while the back page targets timing.

The PF HER integrates these parallel themes into a single coherent system oriented around on-peak consumption, producing a denser but potentially stronger experience for the right audience. Rather than relying on a “Bill NC” data source for each customer as the other variants, an “Hourly NC” renders both aggregate monthly on-peak comparisons and individual average hourly usage comparisons for each bill period. The space for the Benchmark on the front page is replaced with a condensed version of the TOU 101, a “mini” TOU reminder that focuses on weekday rates. The NC module is now emphasizing on-peak usage: the bar chart compares the customer's on-peak kWh against the same peer groups as the other NC modules; the Fair/Good/Great benchmark injunctive norm that reflects that peak-specific status is now displayed as part of the heading of the NC module. The PF NC module’s insight copy adjacent to the 3-bar chart has two options: first, a similar proportional insight anchored on peak usage (“your on-peak usage was 20% lower than similar homes”), and a second configurable option we refer to as the CTG insight that translates the on-peak usage gap directly into annualized dollar savings (e.g., "$120 could be saved per year if your on-peak usage was like efficient neighbors,” activating loss aversion alongside the social-approval mechanism of the injunctive norm (Kahneman and Tversky 1979). On the back page, a new Time of Day Hourly Insight adds a diagnostic layer with no equivalent in either prior design: it displays the customer's average hourly electricity use for the bill period against a comparison group line and identifies the specific peak hour of highest consumption (Oracle n.d.), providing the hour-level precision that converts a norm-induced motivation into a concrete behavioral target (Gollwitzer 1999). The Energy Literacy and Tip module on the back page integrates rate metadata (the on-peak hours) and hourly appliance presence detection algorithms to highlight the appliance categories detected as active during the highest-cost period, ordered by highest likely energy consumption, and pairs that insight with a tip directly tied to one of those categories (Oracle n.d.).

Experimental Design. We tested the Peak Focused HERs at a municipal utility on the west coast with over a half-million residential utility customers. Customers at this utility have been receiving BEE communications since June 2023. All communications use a series of behavioral nudges to encourage customers to modify their energy usage behavior, specifically with the goals of 1) using less energy and 2) using less energy during peak hours (5-8pm).We use an opt-out randomized controlled trial (RCT) design, the standard evaluation methodology for behavioral energy efficiency programs (SEE Action 2012), which ensures statistical equivalence between treatment and control groups on both observable and unobservable characteristics prior to program start; this controls for any additional behavioral treatments live for this utility.

For this study, we set up two RCT experiments: a “legacy” cohort, who has been receiving TOU HERs and 1-2 other Opower communications since June 2023, and a “new” cohort, who has never received BEE communications and met data eligibility requirements to receive HERs over the past year. These eligibility requirements include: an active electric account, at least 12 months of billing observations prior to the BEE program start, a sufficient cohort of neighbors for consumption comparisons, enrollment on a TOU rate plan, and several additional data quality checks. We excluded customers with solar generation, those on medical rates, and other utility-specific criteria prior to identifying the target population.

We identified a legacy population of 253,488 customers and a new population made up of 59,929 customers (see Figure 1). We randomly split the legacy population into three groups: the first treatment group (63,372 customers) received the PF HER with the CTG insight configured, the second treatment group (63,372 customers) received the PF HER without the CTG insight configured, and the third group (126,744 customers) was the Control group (continued to receive the TOU HER experience). We also randomly split the new population into three different groups: the first treatment group (22,587 customers) received the PF HER with the CTG insight, the second treatment group (22,413 customers) received the TOU HER, and the third group (14,929 customers) was Control (did not receive any BEE communication). We created 7 measurements to compare differences in usage between the various treatment and control groups (Table 1).

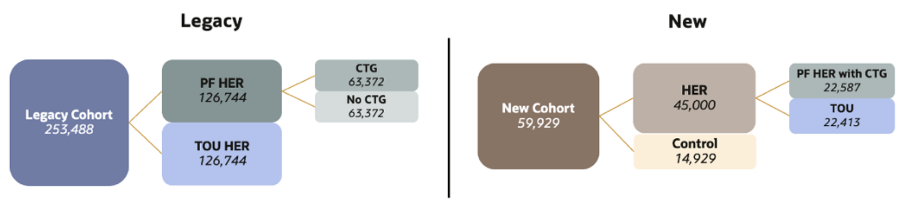

Figure 1. Experimental test design

Table 1: Measurement Names with Treatment and Control Assignments

Measurement Name	Cohort	HER Treatment	Treatment Count	Control	Control Count
(L) PFA vs TOU	Legacy	PF w/ & w/o CTG	126,744	TOU	126,744
(L) PF$ vs TOU	Legacy	PF w/ CTG	63,372	TOU	126,744
(L) PF% vs TOU	Legacy	PF w/o CTG	63,372	TOU	126,744
(L) PF$ vs PF%	Legacy	PF w/o CTG	63,372	PF w/o CTG	63,372
(N) PF$ vs TOU	New	PF w/ CTG	22,587	TOU	22,413
(N) PF$ vs CTL	New	PF w/ CTG	22,587	No HER	14,929
(N) TOU vs CTL	New	TOU	22,413	No HER	14,929

Legend: PF: Peak-focused, TOU: Time-of-use, HER: Home Energy Report, CTG: Close the gap, CTL: Control

Treatment customers began receiving the intervention (experience depending on cohort type) via mail and email in May 2025. The analysis period used for this study is May 1, 2025 through January 31, 2026. We observed both Advanced Metering Infrastructure (AMI, i.e., Smart Meter) hourly interval data and monthly billing records for all customers in the experiment which we used to measure usage differences at the hourly and monthly level between customers in each different experimental group.

Peak kW Savings Model Specification and Calculation. We use a lagged-dependent variable (LDV) regression model specification using customer-level hourly AMI interval data to create estimates of hourly average demand impacts. Our model form is as follows:

$$\begin{aligned} hourly.usage_{i,t} &= \alpha + \beta_1 treatment_i + \gamma_1 avg.pre.same.month_{i,t} + \gamma_2 avg.prior.month_{i,t} \\ &+ \gamma_3 avg.prior.year_{i,t} + \varepsilon_{it} \end{aligned}$$

Where:

- $hourly.usage_{i,t}$ is the electricity consumption, in kWh, consumed by household, *i*, in hour-of-day, *t*,
- $treatment_i$ is an indicator for assignment to the treatment group for household, *i*,
- $avg.pre.same.month_{i,t}$ is an LDV term of the average consumption for household, *i*, in hour-of-day, *t*, in the same calendar month in the year prior to the start of treatment,
- $avg.prior.month_{i,t}$ is an LDV term of the average consumption for household, *i*, in hour-of-day, *t*, in the month immediately prior to the month in which treatment starts,
- $avg.prior.year_{i,t}$ is an LDV term of the average consumption for household, *i*, in hour-of-day, *t*, over the full year prior to the start of treatment and,
- ε_{it} is an idiosyncratic error term.

For each measurement, regression analysis yields the main coefficient of interest, *β*, which is an estimate of the savings per hour per household. We restrict observations included in the estimate (and in the generation of the LDV terms) to exclude weekends, to generate an effect estimate consistent with utility peak reduction goals.

Our approach generates 24 estimates for *β* for each month, representing the average hourly per household demand reduction. We scale these estimates to cohort-level effects by multiplying by the number of customer-days in each month, using the customer account active and inactive dates and the RCT measurement start and end dates.

Monthly kWh Savings Model Specification and Calculation. We use a lagged-dependent variable (LDV) regression model specification using customer-level billing data to create estimates of monthly bill savings impacts. Customers have differing billing periods (bills start and end on different days for each household), so bills can be of non-uniform durations, and may contain algorithmically estimated reads which do not represent actual consumption values. We applied several data cleaning steps to the raw billing data set: estimated reads were coalesced with subsequent actual reads to produce continuous actual-read billing periods; records were calendarized into a uniform calendar-month panel by attributing a weighted daily consumption rate across each month's constituent billing days; and records were trimmed to remove post-move-out usage, daily rates exceeding 300 kWh, and any consumption predating 12 months before program start. Our LDV regression model form is as follows:

$$daily\_usage_{i,t} = \alpha + \beta treatment_i + \gamma \boldsymbol{Y_{o,i}} + mm_t + \left(\gamma \boldsymbol{Y_{o,i}} * mm_t\right) + \varepsilon_{i,t}$$

Where:

- $daily\_usage_{i,t}$ is the average daily consumption for month, *t*, and household, *i,* in the post-treatment period,
- $treatment_i$ is a variable indicating treatment group assignment for household, *i*,
- $\boldsymbol{Y_{o,i}}$ is a vector of three LDV terms, including:
    - *average_preusage$_i$*, the average daily charge for household, *i*, in the 12-months immediately prior to the start of treatment,
    - *average_preusage_winter$_i$*, the average daily charge for household, *i*, in the winter months (December-March) immediately prior to the start of treatment and,
    - *average_preusage_summer$_i$*, the average daily charge for household, *i*, in the summer months (June-September) immediately prior to the start of treatment,
- mm_t is a set of indicators for each month-year, *t*, in the treatment period and,

- $\varepsilon_{i,t}$ is an idiosyncratic error term.

The estimated coefficient, β, is the effect of treatment in the same units as the dependent variable, kWh per household-day. Each observation in the panel is weighted in the regression by the number of days in each month the customer was active, to reflect differing month lengths and customer account closures (e.g., move-outs). Like the hourly model, above, we scale the estimate, β, by the number of customer-days in the analysis period to generate cohort-level effect estimates.

Results

Summary. Peak savings analysis revealed statistically significant savings for 5 out of the 7 measurements during peak hours (Table 2). The three legacy cohort measurements comparing PF HER to the TOU HER showed statistically significant savings during peak hours in at least 3 months. The one legacy cohort measurement that compared those who were configured to receive the CTG insight to those who were not configured did not show statistically significant peak savings in any month. The two new cohort measurements comparing either PF HER or TOU HER to a control group demonstrated the clearest signal during peak hours, with the (N) PF$ vs CTL test resulting in 0.65% savings on average and 10 out of 27 statistically significant peak hours, and the TOU HER versus CTL test resulting in 0.66% savings on average and 11 out of 27 statistically significantly peak hours.

Table 2: Peak Savings Statistical Significance

Measurement	Avg. On-Peak Savings Rate	Max On-Peak Savings Rate	# of Hours with Statistically Significant Peak Savings
(N) TOU vs CTL	0.66%	1.51%	11
(N) PF$ vs TOU	-0.01%	0.29%	0
(N) PF$ vs CTL	0.65%	1.36%	10
(L) PFA vs TOU	0.15%	0.45%	10
(L) PF$ vs TOU	0.18%	0.52%	5
(L) PF$ vs PF%	0.08%	0.24%	0
(L) PF% vs TOU	0.11%	0.44%	6

Average savings rate from 5-8pm, maximum savings rate from 5-8pm, and number of hours with statistically significant saves (out of 27 total peak hours).

Within both the new and legacy cohorts, we did not achieve statistically significant peak savings for any wave until the fourth month of deployment (Figure 2), indicating customers were ramping up their behavior during the initial months. Maximum peak reductions occurred in months with more extreme temperatures (i.e. middle of summer and winter). In the new cohort, the test comparing PF HER and TOU HER did not result in any months of statistically significant savings. In the legacy cohort, this same comparison (PF HER versus TOU HER or rows 4, 5, and 7 in Figure 2) resulted in at least 3 months of statistically significant savings during peak hours.

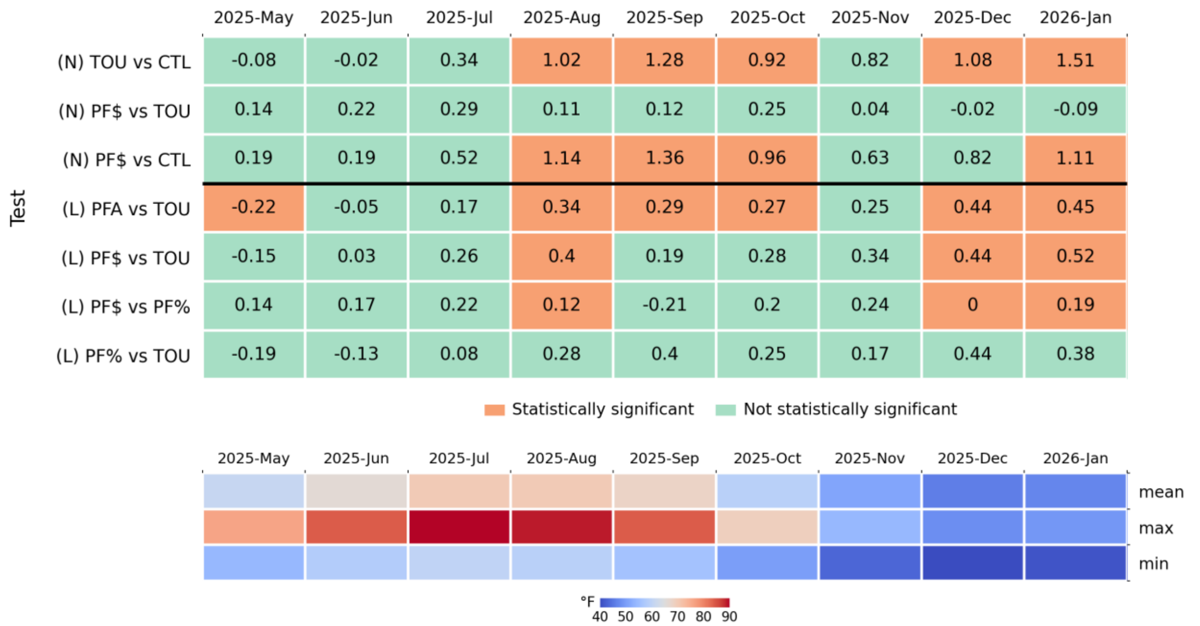

Figure 2: (Top) Maximum on-peak savings rates, coded by statistical significance. (Bottom) Mean, maximum and minimum temperatures, coded by low to high temperature gradient.

Monthly bill analysis did not result in any statistically significant savings. The strongest signal was detected for the new cohort measurements that compared either PF with CTG or the TOU HER to a control group, resulting in 0.26% savings on average during the test period for both measurements; this is consistent with the "ramping period" that typically occurs with HER deployments in the first year. Table 3 outlines monthly and average savings rates by measurement.

Table 3: Monthly Savings Statistical Significance

Measurement	Avg. Monthly Savings Rate	Max Monthly Savings Rate	# of Months with Statistically Significant Savings
(N) TOU vs CTL	0.26%	0.50%	0
(N) PF$ vs TOU	-0.01%	0.02%	0*
(N) PF$ vs CTL	0.26%	0.43%	0*
(L) PFA vs TOU	0.03%	0.19%	0
(L) PF$ vs TOU	0.03%	0.25%	0
(L) PF$ vs PF%	0.00%	0.11%	0
(L) PF% vs TOU	0.03%	0.15%	0

Average monthly savings rate, maximum monthly savings rate, and number of months with statistically significant saves. Most measurements had 9 months of data, * indicates only 7 months of data available.

H1: On-peak consumption between peak-focused and billing normative comparison. The relevant measurements were (N) PF$ vs TOU, (L) PFA vs TOU (L), (L) PF$ vs TOU, and (L) PF% vs TOU. The legacy cohort measurement ((L) PFA vs TOU) resulted in 10 out of 27 peak hours of statistically significant savings (Table 2). The average savings rates during peak range from -0.25% to 0.39%. Three out of the nine months of the experiment resulted in at least one hour of statistically significant savings during peak. The strongest performing month was January (Figure 3), saving at 0.39% on average with the following average savings rates by hour: 5pm: 0.32%, 95% confidence interval (CI) [0.01%. 0.63%], 6pm: 0.40%, 95% CI [0.09%, 0.71%], 7pm: 0.45%, 95% CI [0.14%, 0.76%]. The customers who were exposed to CTG saved at 0.18%, while those who did not get exposed to CTG saved at 0.11%, indicating a slight difference in behavior between the groups, yet not statistically significant. The new cohort measurement ((N) PF$ vs TOU) resulted in no peak savings (Figure 4). The average peak savings rate ranged from -0.34% to 0.22% with no statistically significant hours.

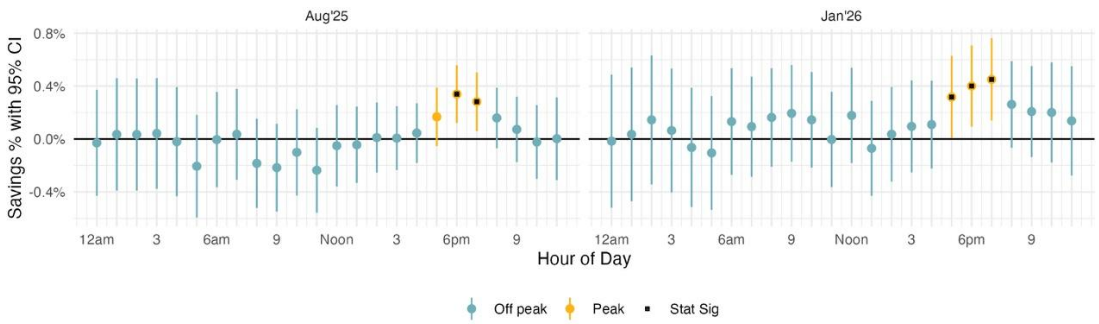

Figure 3: (L) PFA vs TOU test average hourly savings rates in August (left) and January (right) with 95% CI.

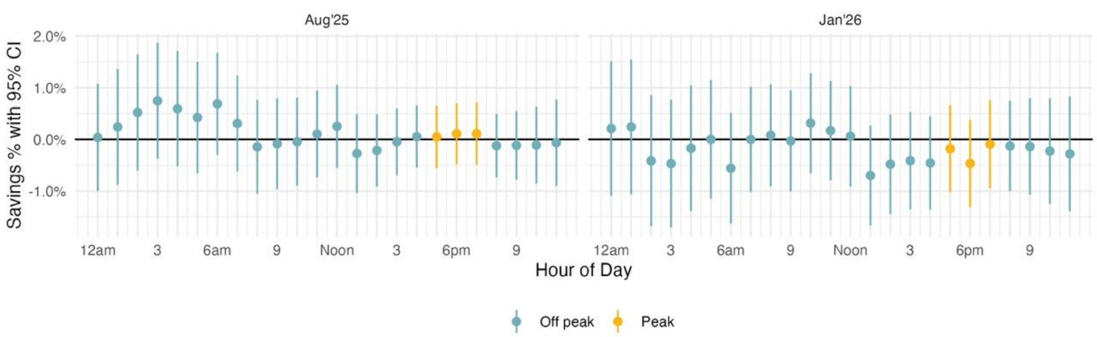

Figure 4: (N) PF$ vs TOU test average hourly savings rates in August (left) and January (right) with 95% CI.

H2: Close the gap Impact on Peak Savings. The relevant measurement was (N) PF$ vs PF%. When compared to the PF HER without CTG, the PF HER with CTG did not yield any statistically significant savings during peak. The average savings rate was 0.08% (Table 2), ranging from -0.22% to 0.19% during peak hours in the test period. July exhibited the strongest savings during peak hours with the following savings rates by hour: 5pm: 0.18%, 95% CI [-0.16%, 0.51%], 6pm: 0.18%, 95% CI [-0.15%, 0.50%], 7pm: 0.22%, 95% CI [-0.11%, 0.55%].

H3: Bill Savings Between Peak-Focused and Standard HER. The relevant measurements were (N) PF$ vs TOU, (L) PFA vs TOU, (L) PF$ vs TOU, and (L) PF% vs TOU. No measurements yielded statistically significant bill savings (Table 3), indicating there is no difference in total monthly energy usage between the varied report experiences. The average monthly savings rates were as follows: (N) PF$ vs TOU: -0.10%. (L) PFA vs TOU (L): 0.03%, PF w/ CTG vs TOU (L): 0.03%, PF w/o CTG vs TOU (L): 0.03%,

H4: Legacy versus new cohort on-peak performance. The relevant measurements were (N) PF$ vs TOU and (L) PF$ vs TOU. While we were unable to set up a true RCT design to compare the on-peak savings between the new and legacy cohorts, we can compare the results of the (N) PF$ vs TOU and the (L) PF$ vs TOU measurements to understand how these two cohorts responded differently within our experiment (Figures 3 and 4). The average savings rate of (N) PF$ vs TOU was -0.01% with a maximum savings rate of 0.29% (95% CI [-0.34%, 0.91%]), which occurred during the 5pm hour of July 2025. The average savings rate of (L) PF$ vs TOU was 0.18% with a maximum savings rate of 0.52% (95% CI [0.14%, 0.90%]), which occurred during the 7pm hour of January 2026. (N) PF$ vs TOU resulted in no months of statistically significant savings, while the (L) PF$ vs TOU measurement resulted in 10/27 hours of statistically significant savings (Table 2).

Qualifying CTG insight firing rate. To supplement the peak and bill savings analysis, we also analyzed the total number of customers that received CTG out of the total treatment population that was configured to receive the CTG insight; customers in the great state, by design, never see the close the gap insight. Across the entire population, only 26% of email (25% of print) recipients in the PF NC fair state and 16% of email recipients (8% of print) in the good state actually saw the CTG insight. Specifically in the study arm examining CTG vs No CTG (the intent to treat), 27% of email (31% of print) reports allowed to show CTG in the study design generated within the good and fair states that could see the actual gap in financial terms. This implies a relative under-firing of the insight based on the test structure and analysis, but reflects the guardrail set up in the module to avoid a crowding-out signal by only showing a gap that equated to greater than $75/year in additional peak energy costs.

Discussion and Future Work

The dominant effect is program-naive customer responses to HER. Among New waves, the standard billing-based TOU HER produced the largest peak reductions: statistically significant savings beginning in August (1.02%) and growing through January (1.51%), with a September peak of 1.28%, in line with the literature on HER (e.g. Allcott 2011; Allcott and Rogers 2014; Munson 2022), but it is occurring specifically on-peak, not in aggregate consumption. The New waves result may represent the high-water mark of a similar trajectory: a strong initial response among customers for whom the normative comparison is genuinely new information. Whether these savings ramp further through the first year or persist into year two and beyond is the most consequential open question this study raises.

The peak-focused content adds modest, accumulating lift among Legacy customers. In the Legacy waves, where customers have been receiving standard TOU HER since 2024, the peak-focused arms show a slightly different pattern: statistically significant effects of up to 0.4% during summer, then statistically significant gains in December and January of 0.44–0.52% (PF% vs TOU and PF$ vs TOU respectively). The gradual accumulation — rather than a spike at the first exposure — is inconsistent with a novelty response and more consistent with habit formation. Customers who have already responded to total-consumption norms may need repeated peak-specific feedback before the new social category (on-peak consumption as a distinct behavioral domain) becomes salient enough to redirect effort. Cialdini, Reno, and Kallgren's (1990) focus theory predicts this: norm salience is a precondition for normative influence, and salience builds with repeated exposure (which happens to be the design rationale for including rate education prominently in the TOU HER and reinforced in PF HER). The growing late-season effects in the Legacy Waves suggest the mechanism is real, but slow to activate in populations already habituated to receiving HER.

The "Close the Gap" dollar framing did not add impact, but it wasn't shown enough. The cleanest test of the loss aversion hypothesis (the Legacy PF$ vs PF% test) is never statistically significant across any month, ranging from −0.21% to +0.24% with no seasonal pattern. This result warrants careful interpretation before it is read as evidence against prospect theory in this context (Kahneman and Tversky 1979). Several attenuating mechanisms are plausible. First, the business logic associated with the "close the gap" message may have led to an under-triggering at report generation time. Upon inspecting the end-states of all customers in the CTG groups, generally fewer than 30% of these recipients received the CTG state. This likely happened because the difference between recipients' peak consumption and neighbor values was not large enough to equate to more than $75 per year of extra cost (the set threshold for this experiment). Second, mental accounting research (Thaler 1999; Hahnel et al. 2020) suggests that making peak costs explicit only opens a distinct mental account for peak expenditure when the dollar amount is large enough to register as a meaningful budget category; this was a guardrail in the product to avoid causing crowding out. At typical residential TOU rate differentials, the implied dollar gap may be too small to cross that threshold for most households, which may suggest the price signal is not strong enough in this market. Third, the study may simply lack power to detect a differential this small against the backdrop of the PF% design's effect itself.

The null result on CTG framing is important for program design: it cautions adding cost-framing complexity to peak HER on the assumption that loss aversion will amplify social norms. Considering how the HER logic was deployed, the peak-comparison mechanism appears to stand on its own; the dollar overlay does not reliably improve on it at scale yet, so more investigation is warranted.

No significant monthly effect. Despite appearing like a problem for the bill period HER, the test structure implies that customers who experienced PF HER in any way experienced similar energy savings to those receiving standard NC in the TOU HER. This suggests that as the new waves ramp up, similar savings results as the TOU HER population in legacy waves can be expected, which continue to show savings on par with most HER programs. Based on this, our interpretation is that the peak-focused savings observed in this analysis are true conservation behavior with an emphasis on peak savings and not load shifting.

Practical implications for utilities. First, for customers already receiving standard HER, the case for switching to peak-focused content rests on a long-run accumulation argument, not an immediate lift: modest but growing effects in the Legacy waves suggest a worthwhile modification, but program managers should not expect a spike immediately after the first mailing. Like most HER programs, expect a ramping up; at this point in the study, we are still likely seeing ramping behavior. Second, the CTG cost framing should be considered carefully. However, adjusting the business logic to fire CTG more often may reveal that actual energy and cost savings during the peak are modest, despite accumulating to significant grid impacts when operating at HER-scale across thousands of homes. Third, the seasonal aspects of on-peak reactions may point to optimal times of year to switch between using the PF signal or adding back the standard NC module in standard and TOU HER.

Limitations. The reported study window runs from May 2025 through January 2026, capturing a summer and partial fall-winter period. The absence of a full winter observation window for the Legacy Waves limits conclusions about seasonal heterogeneity in peak reduction. Effects in July–August for Legacy customers trend positive but do not reach significance; whether a full summer of exposure would change that pattern is unanswerable from this data alone. This study also compared TOU HER recipients to PF HER recipients, primarily receiving email more so than print, because at the launch of their program a few years ago, the utility elected to replace all standard HER deployments with TOU HER and emphasize email communications over print. Future tests should include comparing the HER using NC without any rate education and try to further expand the use of print communications. Generalizability beyond a California municipal utility context—with its specific TOU rate structure, climate, and customer demographics—requires replication in other service territories and could benefit from demographic-sensitive slices to understand if certain populations may respond more strongly. Additionally, this analysis did not dig deeper into the different effects of NC end state and whether CTG was triggered for the groups allowed to receive it. These test results describe the effect on the intent to treat with CTG, but it is not clear if the group that saw CTG would experience more peak savings. Furthermore, CTG was only tested within the legacy arm of the test, not the new cohort. Future work should further unpack the effects of CTG based on specific end-state energy savings, explorations of the CTG triggering mechanism, and new versus legacy customers.

Conclusion

In this paper, we sought to investigate the effects of peak-focused normative comparisons and associated insights. Our findings suggest a value proposition for behavioral load flexibility programs that extend beyond immediate peak demand reduction. Peak-Focused HER addresses these challenges by making peak-period consumption salient through social comparison, highlighting actions likely to impact their peak consumption (which also influence overall usage), and in some cases make the financial consequences concrete through cost framing. For utilities with existing HER programs serving hundreds of thousands of customers, this approach offers a potentially scalable pathway to support load flexibility objectives within an existing HER population and established utility funding mechanism. The intervention leverages the same distribution channels (mail and email) and analytical infrastructure (smart meter data, peer grouping algorithms) already in place for standard billing-based HER, making it a relatively low-cost enhancement to existing behavioral portfolios that appears to have potential for significant impact during peak periods.

The fundamental result here is when a utility describes to customers what their peer group uses during peak hours, what that difference costs the customer, and how the customer should change, it ultimately influences behavior during the target hours in a way that is compatible with energy efficiency conservation efforts. That is a simple design insight with meaningful grid-reliability implications, and one that—at least among utilities that have already invested significantly in AMI and time-varying rate designs—does not require new infrastructure to deploy.

Appendix

Table 4: Acronyms

Acronym	Description
AMI	Advanced Metering Infrastructure
BDR	Behavioral Demand Response
BEE	Behavioral energy efficiency
BLS	Behavioral Load Shaping
CI	Confidence interval
CTG	Close the gap (dollar cost-framing insight)
CTL	Control
DR	Demand response
DSM	Demand-side management
HER / HERs	Home Energy Report(s)
kW	Kilowatt
kWh	Kilowatt-hour
LDV	Lagged-dependent variable
NC / NCs	Normative comparison(s)
PF	Peak-Focused HER
PF%	Peak-Focused HER arm without CTG
PF$	Peak-Focused HER arm with CTG
PFA	Peak-Focused HER arms combined, with and without CTG
RCT	Randomized controlled trial
SEE Action	State and Local Energy Efficiency Action Network
TOU	Time-of-use
TVR / TVRs	Time-varying rate(s)

References

Allcott, H. 2011. "Social Norms and Energy Conservation." Journal of Public Economics 95 (9-10): 1082-1095. 10.1016/j.jpubeco.2011.03.003.

———, and T. Rogers 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." American Economic Review 104 (10): 3003-3037. 10.1257/aer.104.10.3003.

Attari, S. Z., M. L. DeKay, C. I. Davidson, and W. Bruine de Bruin 2010. "Public Perceptions of Energy Consumption and Savings." Proceedings of the National Academy of Sciences 107 (37): 16054-16059. 10.1073/pnas.1001509107.

Cialdini, R. B., R. R. Reno, and C. A. Kallgren 1990. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." Journal of Personality and Social Psychology 58 (6): 1015-1026. 10.1037/0022-3514.58.6.1015.

Faruqui, A., and S. Sergici 2010. "Household Response to Dynamic Pricing of Electricity: A Survey of 15 Experiments." Journal of Regulatory Economics 38 (2): 193-225. 10.1007/s11149-010-9127-y.

Fischer, C. 2008. "Feedback on Household Electricity Consumption: A Tool for Saving Energy?." Energy Efficiency 1 (1): 79-104. 10.1007/s12053-008-9009-7.

Fredman, D., E. Palchak, C. Koliba, and A. Zia 2018. "Not So Fast: The Nuanced Benefits and Risks of Real-time Feedback, Incentives, and Demand Response in Rental Households." Proceedings of the 2018 ACEEE Summer Study on Energy Efficiency in Buildings 8:1-8:12. Washington, DC: ACEEE.

Guidehouse Inc. 2025. “Final Annual Report to the Pennsylvania Public Utility Commission (PA PUC): Program Year 16.” https://www.puc.pa.gov/pcdocs/1896291.pdf.

Hahnel, U. J. J., C. Ortmann, L. Korcaj, H. Spada, and T. Brosch 2020. "Mental Accounting Mechanisms in Energy Decision-Making and Behaviour." Nature Energy 5 (12): 952-958. 10.1038/s41560-020-00704-6.

Kahneman, D., and A. Tversky 1979. "Prospect Theory: An Analysis of Decision under Risk." Econometrica 47 (2): 263-291. 10.2307/1914185.

Karlin, B., J. F. Zinger, and R. Ford 2015. "The Effects of Feedback on Energy Conservation: A Meta-Analysis." Psychological Bulletin 141 (6): 1205-1227. 10.1037/a0039650.

Kesternich, M., A. Lange, and B. Sturm 2019. "Pro-environmental Incentives and Loss Aversion: A Field Experiment on Electricity Saving Behavior." Energy Policy 137: 111131.

Munson, T. 2022. "Testing the Efficacy of Home Energy Reports." ILLUME Advising, LLC. https://illumeadvising.com/2022/case-study-testing-the-efficacy-of-hers/

Murphy, S., C. Miller, and J. Deason 2021. The State of Demand Flexibility Programs and Rates. Lawrence Berkeley National Laboratory. CEC-500-2021-024.

Nicolson, M. L., M. J. Fell, and G. M. Huebner 2018. "Consumer Demand for Time of Use Electricity Tariffs: A Systematized Review of the Empirical Evidence." Renewable and Sustainable Energy Reviews 97: 276-289. 10.1016/j.rser.2018.08.040.

Oracle. n.d. "Oracle Energy Efficiency Cloud Service Documentation." Accessed February 7, 2026. https://docs.oracle.com/en/industries/energy-water/energy-efficiency/index.html

SEE Action (2012). "Energy Efficiency Program Impact Evaluation Guide." State and Local Energy Efficiency Action Network. U.S. Department of Energy.

Sloot, D., N. Lehmann, A. Ardone, and W. Fichtner 2023. "A Behavioral Science Perspective on Consumers' Engagement With Demand Response Programs." Energy Research Letters 4 (1). 10.46557/001c.38831.

Thaler, R. H. 1999. "Mental Accounting Matters." Journal of Behavioral Decision Making 12 (3): 183-206. 10.1002/(SICI)1099-0771(199909)12:3<183::AID-BDM318>3.0.CO;2-F.

Vine, D., L. Buys, and P. Morris 2013. "The Effectiveness of Energy Feedback for Conservation and Peak Demand: A Literature Review." Open Journal of Energy Efficiency 2 (1): 7-15. 10.4236/OJEE.2013.21002.

Wolak, F. A. 2011. "Do Residential Customers Respond to Hourly Prices? Evidence from a Dynamic Pricing Experiment." American Economic Review 101 (3): 83-87. 10.1257/aer.101.3.83.

Acknowledgments

The authors would like to thank Kelly Greer, Carol Ho, Russell Meyer, and Sanem Kiser and for their support developing the program design, and analysis, assisting with test configuration, and for providing intermediate reviews and constructive feedback during the writing of this paper. We also would like to acknowledge the Opower HER design and engineering team who worked tirelessly to bring this new solution to life to facilitate rigorous test-and-learn capabilities.

AI-assisted tools used: Claude Sonnet 4.x, OpenAI GPT 5.x
Used for: background research and paper preparation/revision.
Extent: partial drafting of introductory sections, light editing, citation formatting.
Human review: The authors reviewed and verified all content, calculations, and citations.