MASTER

Sustainable Goals

Energy-Saving Goals and Signposting in a Rasch-Based Recommender System

Oonk, Aleid S.

Award date:
2023

07-07-2023

Sustainable Goals: Energy-Saving Goals and Signposting in a Rasch-Based Recommender System.

by Aleid Oonk

identity number 0964050

in partial fulfillment of the requirements for the degree of

Master of Science
in Human-Technology Interaction

Supervisors:

dr. ir. M.C. Willemsen, IE&IS, Eindhoven University of Technology
dr. ir. A.D. Starke, Social and Behavioural Sciences, University of Amsterdam

Third Assessor:

Prof. dr. C.C.P. Snijders, IE&IS, Eindhoven University of Technology

Abstract

To aid consumers in energy conservation in a household context, recommender systems can help users find suitable saving measures. Rasch-based energy recommender systems have been successful in providing users with ability-tailored saving recommendations and can present users with items of appropriate difficulty. This ensures an optimal trade-off between novelty and feasibility and works better than a one-size-fits-all system. Several manipulations of such systems have been studied, including the presentation of a fit score next to items, and indications of the percentage of peers who perform certain measures. These manipulations resulted in between-item effects, but no overall effects on savings.

In our system, we asked 202 participants to test and evaluate our 'Saving Aid' Rasch recommender system. The current study examines whether guided goal setting influences user experience, user perceptions, and energy savings in such a system. We furthermore look at the effects of value activation through signposting, for which we presented saving metrics as either kWh, Euro, or CO2 values. We then looked at the interaction of signposts with pro-environmental values (NEP scale) and financial values (IMS scale) of users. Furthermore, half of the participants chose a saving goal. After approximately four weeks, participants were asked which measures they actually performed.

Although we did find that the CO2 signposts resulted in slightly lower chosen savings, we only saw a very small interaction effect with personal values in the initial study, and this did not translate into an effect on actual savings after four weeks. We did observe various interaction effects of the CO2 signpost and NEP scores on user experience variables, e.g., on choice difficulty, choice satisfaction, and energy-saving self-efficacy. We did not see any effects of guided goal setting in the initial study or in the follow-up study after four weeks.

Keywords: Recommender systems, sustainability, personalization, goal setting, signposting, user-centric evaluation, Rasch model, user experience

Acknowledgements

I could not have completed this project without the help of my supervisors, Martijn Willemsen and Alain Starke. I would like to express my sincere gratitude to Martijn for the immense support and guidance over the past half year. I am still very happy you suggested working on this project, and I learned a lot from you, from web development to new statistical tools, how to carry out an online experiment, and overall, to challenge myself during this project. The meetings were always fun, and it was never a problem if they took a little longer than scheduled. You were quick to respond when I asked for help as well. I really could not have asked for better supervision, and I hope you enjoyed working on the project as much as I did.

Alain, thank you for the support and the time you invested in this project as well! It was wonderful to have the person who originally came up with the system help guide this project, and you generously shared your knowledge and experience throughout. From before the project even started, you helped me to get started with the web-development part, and you even attended several meetings on campus despite not working here anymore. Also, big thanks for ensuring that the project remained manageable when it became a little over-ambitious at times. You always asked the right questions about things that otherwise might have been missed, and your attention to detail helped me improve this thesis as well. I very much appreciate your help over the past half year.

And of course, thank you, Chris Snijders, for agreeing to be the third assessor.

I would also briefly like to thank Vaida and Emma for their greatly helpful thesis writing seminar (though they did not encourage this kind of page count), the participants who joined this study for their time and effort, and Jan Tjemme van Wieringen for sharing some valuable insights on household energy usage.

Lastly, I would like to thank my parents, brother, and friends for their support and encouragement throughout my studies, especially during the past half year. I could not have completed this without you.

Aleid

Table of Contents

1. Introduction

For various reasons, interest in energy conservation is on the rise (Domalewska, 2021). On the one hand, increasing energy prices might warrant a reduction in energy consumption, and on the other hand, a reduction in energy use would be desirable from an environmental point of view. To limit current global warming to 1.5 degrees Celsius, emission reductions towards a net zero or below are needed (United Nations [UN], 2022). While this requires cooperation from governments and industries alike, limiting individual and household energy consumption is a substantial part of this puzzle. To aid users in adopting suitable energy-saving measures, recommender systems could be utilized that present users with personalized recommendations on energy-saving measures. Recommender systems are interactive systems that can present information to users that is predicted to be relevant to them. In the past, Rasch-based energy recommender systems have been found to be more effective than one-size-fits-all approaches in the energy-saving domain (Starke et al., 2020). Drawing from research in other domains, this study will look at the effects of goal setting and attribute translation through signposting on savings within such a Rasch-based energy recommender system.

1.1. Outline

In the theoretical background section, I will first outline the current situation regarding climate change, energy conservation, and household energy use. Then, I will cover the basics of recommender systems, their opportunities and limitations, and their application in the energy-saving domain. After, I will explain the basics of Rasch models, Campbell's paradigm, and relevant applications. From there, I will outline previous research on Rasch-based energy recommenders, including implementations of nudges and social norm depictions. Next, I will cover existing theories on goal setting, the application of goal setting in recommender systems, and their role in energy conservation. Lastly, I will discuss the findings obtained in signposting research and how this could be combined with goal-setting approaches to ultimately improve system efficiency.

2. Theoretical Background

2.1. Climate Change

The impact of human-caused climate change is becoming increasingly visible, and the need to limit the environmental damage is becoming increasingly urgent. Atmospheric CO2 levels have risen by 50 percent since 1750 and have exceeded levels of the last 400,000 years (National Aeronautics and Space Administration [NASA], 2023). This increase correlates strongly to the rise in fossil fuel burnings (NASA, 2023; NASA, 2022), and these resulting greenhouse gas levels have caused concerning increases in global temperatures (Callendar, 1938; Kweku et al., 2018, International Energy Association [IEA], 2022). Since the 18th century, the average global temperature has already risen by 1.1°C (Intergovernmental Panel on Climate Change [IPCC], 2021), and sea levels have risen by around 20 centimeters (Lindsey, 2022). Further increases will result in the occurrence of more frequent extreme weather conditions, including hurricanes, heat waves, and droughts, increased flooding of coastal areas, and an increase in marine heatwaves (Dosio et al., 2018; IPCC, 2021; Mousavi et al., 2011; World Health Organization, n.d.). To prevent such worsening scenarios, world leaders from 198 countries and the European Union have agreed to limit this temperature increase to well below 2°C and strive towards a maximum 1.5°C increase compared to the pre-industrial era (Schleussner et al., 2016).

2.2. The Role of Individuals

To restrict global warming to these limits, emissions need to be reduced by 45% by 2030, and reach at most net zero by 2050 (UN, 2022); implying that no more greenhouse gases should be emitted than there is capacity for them to be re-absorbed. While governments and industries can play a role by encouraging and adopting sustainable practices, individuals can also contribute: Household energy use accounted for 29.9% of the total world energy consumption as of 2018 and is increasing at a much faster rate than industrial energy consumption (IEA, 2020; Matsumoto et al., 2021). Therefore, limiting energy consumption in our day-to-day lives is an important factor in reaching this goal. While this thesis will mostly focus on direct energy consumption through for example heating, cooling, and household appliances; ultimately, indirect energy consumption, for example through clothing, food, and travel, should also be considered to reduce the total energy consumption.

2.3. Household Energy Use

To be able to reduce energy consumption, it is important to understand where energy is being used and where it might be used in higher quantities than necessary. For this study, we will consider Dutch households, and the below section also applies to this target group. Note that the energy usage situation might be different in other countries.

According to Milieu Centraal (n.d.), 80% of gas usage in Dutch households is used for heating of rooms, and 20% is used for heating water for other purposes, e.g., for showering. This division varies per dwelling, and each type of dwelling will require different strategies to bring down energy consumption. A large, freestanding house generally has more external heat losses than an apartment with neighbors all around. In a badly isolated house, up to 980 cubic meters of gas can be saved per year by turning down the heating (Milieu Centraal, n.d.). Furthermore, around one in four people could save energy by limiting their heating to 19°C, over half of people could decrease energy consumption by not heating unused rooms, and four in ten people could lower the heating at night to 15°C to save energy (Motivaction and Milieu Centraal, 2023).

In addition to measures that focus on heating, there are countless other energy-saving methods to be considered. For example, consumers could save energy by buying more energy-efficient appliances, installing better insulation, or putting on extra clothing instead of turning up the heating (Bams, 2018; Starke 2019). They could also opt to completely turn off appliances when not in use, install a smart thermostat, or use radiator foil to limit their energy consumption. Because there are so many ways to save energy, each with their own costs, benefits, and drawbacks, getting an overview and making a choice between these measures can be overwhelming.

Energy Saving Measure Dimensions. To get a better overview of energy-saving measures, there are various ways to differentiate between them. For example, Boudet et al. (2016) use a four-dimensional construct to evaluate energy-saving measures, based on the frequency of the measure, the skill needed to perform the measure, the observability of the measure, referring to who can notice the measure having been performed (e.g. only by the person who performed the measure, or also by guests) and the locus of decision, referring to who decides to execute a certain measure. Based on this, they distinguish four clusters: measures that can be performed by all members of the family, referred to as the *'family style cluster'*, an *'expert cluster'* with infrequent, costly measures that require high skill and are less easy to be observed by others, a large cluster of *'management measures',* that are less frequent, but easy to perform and less observable, and the last cluster is the 'weekend project cluster', consisting of low-cost measures that require some amount of skill, but that are not as frequent, such as installing energy-efficient appliances.

Curtailment vs. Efficiency Measures. Another way to distinguish measures is in a two-dimensional way, based on cost and frequency alone, often resulting in two clusters of curtailment and efficiency measures (Karlin et al., 2014). Curtailment measures are low-cost, high-frequency behaviors, e.g., washing at a lower temperature or using a lid during cooking, whereas efficiency measures refer to high-cost, low-frequency behaviors, including changes to dwelling or appliances. Examples of the latter category are installing better insulation, installing low-flow showerheads, or opting for an energy-efficient freezer. In the definitions of curtailment and efficiency measures, cost can refer to both behavioral and financial costs. Several measures are considered both low frequency and low cost, e.g., replacing light bulbs. These might get categorized as maintenance measures (Karlin et al., 2014). We will for now only consider the curtailment-efficiency distinction.

Although efficiency measures often result in higher savings, many people prefer curtailment measures (Lesic et al., 2018). In part, this might be due to financial restrictions: Individuals with a higher income are more likely to perform efficiency measures, and less likely to perform curtailment measures (Umit et al., 2019). Moreover, there are differences in these behaviors between citizens from different countries (Umit et al., 2019). Therefore, there is room for improvements, specifics of which might depend on the exact target population.

Rasch Scale. In the rest of this report, rather than using the four or two-dimensional construct described above, we will distinguish measures based on a one-dimensional Rasch scale that considers only a difficulty dimension. This makes it easier to match people to measures, as you also only have to evaluate people on one dimension (in this case, skill). We will further explain the workings of the Rasch scale in section 2.6.1. *'Rasch-based energy recommender systems'*.

Whichever distinction one uses to make effective recommendations, these recommendations are only the start: they will only be effective when people are motivated to implement the recommended measures. Therefore, it might be useful to consider behavioral trends concerning energy consumption and see if, and how much, room for improvement there is.

Consumers do seem to be motivated to reduce their energy consumption: Between 2010 and 2018, gas usage already decreased by 20% (Milieu Centraal, n.d.), and in 2023, around 70% of Dutch consumers lowered their thermostat as compared to the previous year, with an average of a 2°C (Motivaction and Milieu Centraal, 2023). Most people lowered the temperature due to increased energy prices, and one in six households experienced difficulties with paying energy bills. People were furthermore motivated by the environment and a desire to be less dependent on gas. 80% of people with a boiler indicate that they intend to continue to save energy, even though a gas and electricity 'price ceiling' was implemented by the Dutch government. Therefore, an energy-saving recommender system might still be relevant and helpful to consumers, and increasing the effectiveness of such a system could be beneficial both from an environmental viewpoint as well as from a consumer viewpoint.

2.4. Attitude-Behavior Gap

Although consumers often recognize the importance of protecting our environment and might also explicitly claim to be concerned with the environment, acting on such values is not always as straightforward, despite efforts being made. The discrepancy between environmental attitudes and corresponding pro-environmental behaviors is commonly referred to as the attitude-behavior gap or value-action gap. This gap has been observed in sustainable product choices, green travel choices, and in household energy consumption (Newton & Meyer, 2013; Zhang et al., 2021). Several causes of this gap can be identified, including experienced difficulties in finding relevant information, difficulties in working out what would be the best course of action, and time and financial constraints (Newton & Meyer, 2013). Therefore, merely considering the extent to which someone has pro-environmental values, is not enough: it is important to consider that pro-environmental values might not automatically translate into pro-environmental actions, and it is crucial to identify and reduce possible obstacles that people might face in translating their values into actions.

2.5. Recommender Systems

To help consumers overcome the barriers posed by choice difficulties and information deficiencies, recommender systems could be of added value. These are systems that present users with items that are predicted to be relevant to them (Lü et al., 2012). They can be used to make sense of large quantities of data and a multitude of options (Lü et al., 2012), and are employed in various domains ranging from music (e.g Spotify) movies, series (Netflix), and travel, to health, nutrition, and fitness applications. They can employ a one-size-fits-all approach, or personalize recommendations based on user characteristics, preferences, and/or behavior. A non-personalized algorithm might recommend the most popular items, commonly referred to as a top-N algorithm, or present items that are sorted on a certain characteristic, e.g., kWh savings. A tailored system might make predictions by drawing on data of similar users (user-user), by recommending similar items to those previously preferred (item-item), or by evaluating patterns of behavior through collaborative filtering (Koren & Rendle, 2021).

What works best might depend, among others, on the user, use, and context of the system. While recommender systems provide many opportunities, they are also accompanied by various challenges; for example, tailored systems need a certain amount of user data to draw from (cold start problem), preferences of users might change over time, and tradeoffs between diversity and accuracy must be made to optimize these systems (Lü et al., 2012). Additionally, user interfaces might influence the effectiveness of recommender systems (Lü et al., 2012).

2.6. Energy Recommender Systems

Recommender systems have frequently been used in the energy domain, though often based around intelligent energy systems (Himeur et al., 2021). To use such a system, one must often have a smart meter installed. Given that some systems draw upon data from various rooms and outdoor sensors, this might further increase the costs of implementation. An app-based (Rasch) recommender system requires no physical installation and might be more readily accessible to a larger group of people.

2.6.1. Rasch-Based Energy Recommender Systems

Starke et al. (2020) created an online Rasch-based energy recommender to present tailored energy-saving measures without a need for elaborate energy usage data from users. This system was found to be more efficient than a similar one-size-fits-all system. This Rasch system is based on Campbell's paradigm and compares both item difficulty and user ability on the same one-dimensional scale, to predict adoption probabilities of certain items by certain users. Difficulty on a Rasch scale is related to the proportion of people who perform a certain item X, and ability is based on the proportion of items a certain individual performs. The probability *P* that an individual *n* performs a certain measure *i* can be calculated using the following formula (1):

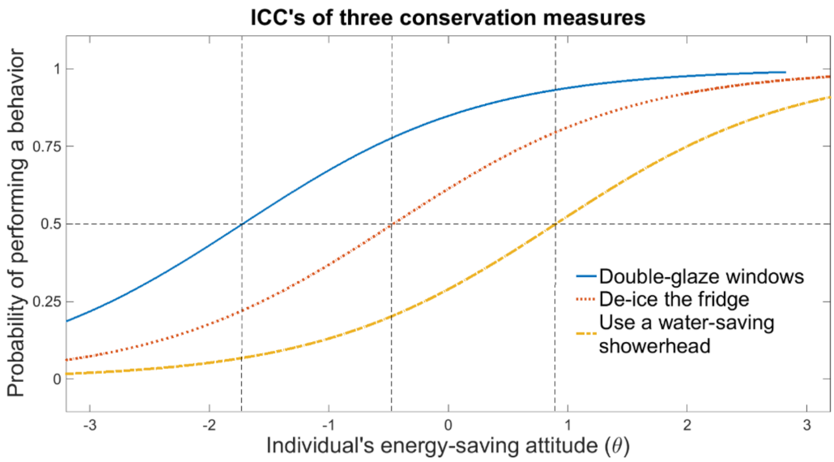
$$P\{X_{ni} = 1\} = \frac{e^{\theta_n - \delta_i}}{1 + e^{\theta_n - \delta_i}} \qquad (1)$$

In which θ is the individuals' ability, and δ is the measure's behavioral costs. The behavioral cost of a measure is equal to the ability level that results in a 50% engagement probability with that measure. In which case, exponents in the above formula become 0.

This formula results in item-characteristic curves (ICC's) that depend on item difficulty, as can be seen in Figure 1; indicating the likelihood (Y-axis) that an individual with a certain ability/attitude (X-axis), performs a certain item (different curves). An easier item will thus result in a curve with higher likelihoods ratios (e.g., the blue curve in Figure 1, depicting the likelihood *Y* that someone with ability *X* with install double-glazed windows) than a difficult item (e.g., the yellow curve in Figure 1, depicting the likelihood *Y* that someone with ability *X* will use a water-saving showerhead). Difficulty in a Rasch-sense is therefore directly related to the popularity of a certain measure, and not necessarily to the objective cost or effort that would be needed to perform a measure.

Figure 1

Item characteristic curves of three energy-saving measures, obtained from Starke et al. (2020)



In the study by Starke et al. (2020), users indicated on a list of thirteen energy-saving items of systematically varied difficulty, which ones they already performed. From this, an ability score between 0 and 13 was computed, which was then used for presenting tailored recommendations. The underlying assumption here is that users with a certain ability *X*, are more likely to also perform items with a difficulty of *X* or below, than they are to perform an item with a difficulty of above *X*.

Tailoring measures in this way can optimize the trade-off between on the one hand presenting novel measures that a user not yet performs, and on the other hand presenting feasible measures that are within a user's ability. It was found that users preferred items slightly below their ability. While the difference in efficiency was not outstanding compared to indiscriminately recommending items of average difficulty, personalizing a system to fit user characteristics and preferences, might still improve user experience as observed in other studies

(Knijnenburg et al., 2011). When looking at research in other applications of recommender systems, there might be room for further improvements of this system. Before I suggest such opportunities, I will first discuss several adaptations to the system that have already been explored.

2.6.2. Nudging in Rasch Energy Recommenders

Starke et al. (2017) found that tailoring recommendations and highlighting certain items in those personalized lists with a star, reduced perceived and actual effort in users. This manipulation made users feel more supported by the system, increased choice satisfaction, and increased the total number of measures chosen. These effects furthermore depended on the order of difficulty of the lists (ascending or descending). It was also observed that the increased level of perceived support for a tailored system was inversely correlated to user ability: users with higher ability might have had less need for this kind of assistance. While more measures were chosen for a tailored system, these tended to be of lesser difficulty, and when surveyed four weeks later, users were not more likely to perform those measures. It was furthermore found that users were more likely to perform the easier measures than the more difficult ones.

Fit scores. Starke et al. (2017) additionally studied the effect of presenting a ‘fit’ score next to measures, a percentage indicating the appropriateness of a measure, based on the Rasch scale. This score was mostly above 60% with a median of 77%. This study had 3 difficulty conditions: Relatively easy, matching difficulty, or relatively difficult recommendations. The effect of this score differed between conditions: users in the relatively easy condition picked more items, while users in the relatively difficult condition picked fewer items and were less satisfied with the system when presented with a fit score. The authors state that fit scores therefore merely reinforced current ability levels.

Item Characteristics. In his master thesis, Bams (2018) found that users are unaffected by kWh ratings of items but did tend to choose items with lower perceived and objective effort ratings. Nudging users by showing saving scores, and smart saving scores (corrected for effort needed), only had a within-list effect: items were more likely to be chosen, at the expense of other items.

Because these scores did not direct users to choose items with higher kWh savings, there was no cumulative effect on total kWh saved.

Social Norms. Starke et al. (2021) studied the effect of depicting social norms in the Rasch energy recommender. Energy measurement items were accompanied by percentages indicating the proportion of people performing a certain action. These were framed as the proportion of people with either lower, similar, or higher experience than the user, or as a global average. They found an effect of depicting social norms, but again, only within lists of recommendations;

Higher descriptive norm percentages resulted in higher likelihoods of items being chosen, at the expense of other items. As these items were not necessarily higher kWh, there was no net effect on total kWh saved, nor was there an effect on the number of items chosen. The same within-list effect was not observed for a control condition depicting saving scores of items: more efficient items were not more likely to be chosen for this presentation. For the similar and global descriptive norms condition, there was an indirect positive effect on choice satisfaction, which was mediated by higher perceived feasibility. Furthermore, users with stronger concerns about their role in protecting the environment, rated items as more feasible to perform.

While the actual effects on energy savings of the three previous manipulations are somewhat unclear, there are indications that these kinds of nudges can influence user satisfaction. For example, Starke et al. (2021) found a positive effect of social norm depictions on perceived feasibility and choice satisfaction despite not finding an increase in total savings. Increased satisfaction of users with the system might in turn help encourage repeated use of the system. Therefore, such manipulations might still contribute to eventual energy savings in the long run.

2.7. Goal Setting

In previous Rasch-based energy recommender studies, individuals were not encouraged to adopt a specific number of measures or reach a certain number of kWh savings. This might explain why previous manipulations ultimately did not result in higher savings. It might be that certain manipulations are more effective when users are encouraged to reach a certain goal; especially for users who might not be intrinsically motivated to reduce their energy consumption.

2.7.1. Goal-Setting Theory

In their goal-setting theory, Locke and Latham (2002) state that the mere presence of goals increases performance by guiding efforts in a desired direction. Goals that are neither too easy nor too difficult lead to the highest satisfaction, and they should be of limited complexity. Below, I will outline several characteristics of goals, values, and their effects on performance outcomes.

Goal Level. Ryan (1970), (as cited in Locke and Latham (1991)), hypothesized that interpersonal differences in task performances are caused primarily by differences in individuals' own performance goals. Locke and Latham (1991), extend this theory by exploring the links of goal setting, ability, commitment, and motivation. They observed that higher goals lead to higher performance, as long as there is sufficient ability and sufficient commitment to reach these goals. When commitment is lacking, the height of the goal has little effect on performance levels.

Values. This strongly relates to what Eccles and Wigfield (2002) describe in their literature review on motivational beliefs, values, and goals. They describe the difference between intrinsic and utility values of actions. Intrinsic values describe, among others, the enjoyment an individual experiences when performing an action. The utility value of an action, on the other hand, describes the extent to which a task aligns with an individual's goals. For the energy-saving domain, the tasks themselves might not be particularly enjoyable (turning off a monitor or letting someone else install solar panels); However, these tasks could contribute towards achieving personal goals, like behaving in a more sustainable way, or saving money.

Goal Frames. Another way to understand the link between values and goals is through the Lindenberg and Steg (2007) theory on goal framing in the environmental behavior domain. They distinguish three goal frames: The *hedonic* goal frame, which views goals from the perspective to which they help an individual feel better in a situation. This hedonic perspective is similar to the intrinsic value definition described by Eccles and Wigfield (2002). Lindenberg and Steg also describe a *gain* goal frame, which focuses on acquiring resources. This relates to the utility values of actions described by Eccles and Wigfield (2002). Lastly, they describe the *normative* goal frame, in which actions are neither evaluated through a lens of immediate enjoyment nor on their contribution towards personal gain, but rather through a lens of what ought to be the right course of action in a moral sense.

For energy saving, people might be motivated by both personal gains, as well as by a sense of duty or responsibility towards the environment or society as a whole. The design of a system might influence how saving goals are perceived (whether they contribute to personal gain, societal benefit, or immediate enjoyment), and depending on existing user values, a system might be more or less effective. I will further elaborate on this in the section on signposting.

Goal Source. In addition to how the contents of a certain goal are perceived, the source from which a goal originated might also impact willingness to act, as well as influence beliefs in one's ability to achieve a goal. Locke and Latham (1991), across multiple studies, observed that goals assigned by an authority figure tend to carry over to other situations: When people consequently choose their own goals in similar contexts, they do so in line with the authority's directions. This might be in part due to the commitment people experience from receiving such a task, but also by a belief that the authority knows what a reasonable goal would be (Locke and Latham, 1991). An assigned goal was found to increase self-efficacy in people who now believed they could in fact reach this goal. Therefore, some guidance on what appropriate goals would be from the system (given that people view the system or the people who made the system as some kind of authority), might help motivate people to reach these goals.

Performance Feedback and Adjustment. Often, goals are no one-time occurrences. Behavior changes and outcomes might be a longer-term endeavor, and thus feedback on progress and performance helps users evaluate and adjust their performance. Feedback is an important aspect of goal setting: people need an indication of how their performance compares against their goal to be able to adjust their performance to be in line with that goal (Locke & Latham, 2002).

In previous research, it was found that, although goal setting by itself increases performance, formal feedback further amplified this effect (Kim & Hamner, 1976). Kim and Hamner found that feedback was more effective if it came from an outside source, as compared to letting users evaluate themselves (referred to as an 'intrinsic' source of feedback). Feedback was also more effective when it was given in a more formal setting as opposed to during an informal conversation. Therefore, presenting users with a reasonable goal, and allowing users to see how their actions lead towards achieving that goal, might further encourage users to save energy, because user efforts would be directed towards a certain amount of savings.

Goal Complexity. Wood, Mento, and Locke (1987) concluded from their literature review that the effect of goal setting is less pronounced on more complex tasks than it is on simple tasks. Locke and Latham (2002) state that this might be due to peoples' limited ability to think of strategies for reaching a goal; The more difficult a goal is, the harder it is to think of appropriate strategies and thus the harder it is to attain a goal. A single goal might be experienced as less complex than a whole list of mini-goals or recommendations. A single goal could provide a clear focal point and serve as a reference to compare current behavior to, whereas a list of multiple actions might be more difficult to maintain a mental representation of.

To summarize this thesis so far; Rasch energy-saving recommender systems were found to be more effective than one-size-fits-all solutions, and several nudges were found to have a within-list effect but did not result in higher total savings. From goal setting literature, it becomes apparent that goal setting might be a potential way to direct user actions towards higher energy savings, and guided goals might ensure that goals are realistic for users. The effect of goal setting on savings and user satisfaction will therefore be examined in this study.

[RQ]: *What is the effect of (guided) goal setting on choice satisfaction, self-efficacy, and energy savings in a Rasch-based energy recommender system?*

Locke and Latham (2002) do mention as a limitation of their theory that goal conflict might be a barrier for its implementation: when people have conflicting goals, performance might be undermined. In the case of energy savings, financial objectives or constraints might conflict with energy-saving goals. This might undermine pursuits to implement certain measures. However, in a Rasch-based system, these kinds of difficulties generally result in a higher item difficulty rating, thereby this trade-off is accounted for in an indirect way.

2.7.2. Goal Setting in Energy-Saving Applications

Goal setting in the broader energy-saving domain has been studied with mixed results. Several aspects of goal setting, including the degree of autonomy, the extent to which goals are realistic pursuits, and previous performance outcomes, might affect user motivation, commitment, and total energy savings.

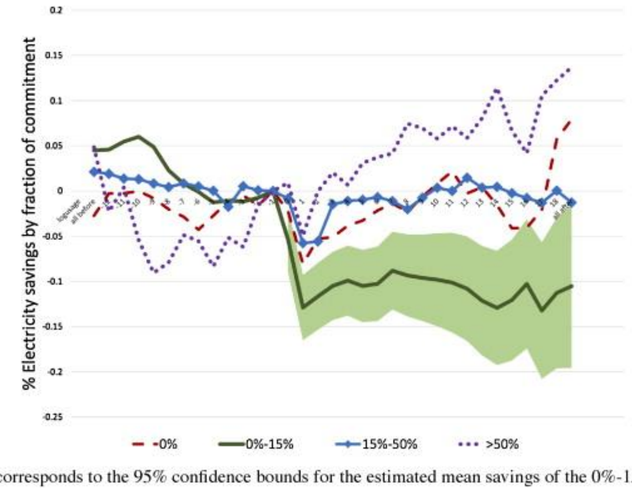
In a study by Abrahamse et al. (2007) on energy usage in the province of Groningen of the Netherlands, an effect of goal setting and education on energy consumption was found, albeit not significant due to large within-group variability. However, when indirect energy use through travel movements was excluded, the effect was significant. In this study, participants filled out an extensive list indicating which household appliances they owned, and how often they used them. Additionally, indirect energy consumption through food and travel was taken into account. Based on this data, users were presented with tailored energy-saving advice.

It was found that instructing households to reduce their energy consumption by 5%, along with performance feedback and education on energy usage, resulted in an energy reduction of 5.1% on average over the course of a 5-month study. A stronger effect was seen in the group that got feedback on group progress in addition to their own progress. While the baseline energy consumption was based on the previous year's energy consumption data, measured reductions were based on self-reported behavioral changes. Therefore, social desirability bias might have been at play in this study. Nevertheless, this study indicates that goal setting could be effective within an energy-saving recommender that targets direct energy consumption, as opposed to indirect consumption.

Self-Assigned Goals. In a study by Harding and Hsiaw (2014), users were asked to select energy-saving measures from a list of recommendations. They found that the level of the self-assigned energy reduction goals greatly impacted eventual savings. Users who chose a reduction goal between 0% and 15%, as compared to their current energy usage, achieved the highest energy reductions at around 11%. Users who did not choose any items only achieved a 1.5% reduction, and users who set overly optimistic goals above 15%, tended to give up completely and scored even worse with 1% reductions on average (as seen in Figure 2 which was obtained from this study). They also observed that users tended to over-commit, with 40% of users opting for an overly optimistic goal between 15 and 50%, and 12% of users opting for an 'undoubtedly unreasonable' goal of above 50% energy reductions. This gives further support for the notion that it might be useful to restrict the level of these goals, or to assign goals for users rather than allowing self-assignment. It should be noted, however, that the authors observed a correlation between certain user characteristics and the tendency to over-commit to goals, e.g., indications of problems with self-control. These characteristics might influence the ability to follow through on goals regardless of their difficulty, and thus this effect may or may not resolve when overly optimistic goal setting is restricted.

Figure 2

Energy usage difference between groups, divided by self-assigned goals, obtained from the study by Harding and Hsiaw (2014)



Shaded area corresponds to the 95% confidence bounds for the estimated mean savings of the 0%-15% goals group.

Note: 0 depicts the month that the study started

The Effect of Not Reaching Goals. Fraser (2023) studied 10 years of historical data on participation in an energy-saving challenge in British Columbia, Canada. A utility company serving 1.7 million people encouraged households to participate in a program that promised a $75 reward for a 10% (weatheradjusted) reduction in energy consumption. Although there is no mention of monetary costs to participate in the program, in the first year of the program, only 1.2% of customers participated, and only 60% decided to re-enroll for another year. It should be noted that each additional year, it becomes more difficult for users to reach their goal if they want to continuously reduce their energy usage.

Participants could get feedback on their progress and receive advice to reduce energy use in an online portal. Fraser (2023) studied the anonymous data of 10,000 participants and 20,000 non-participants in total. Participant decisions to either terminate or continue participation were strongly correlated to whether they successfully achieved previous goals. This was irrespective of the degree of success: Just being short of reaching a goal, resulted in an equal likelihood of termination as being far from reaching a goal. It was also observed that the longer households participated, the more energy they saved. However, as soon as they terminated participation, a partial rebound effect was observed with energy consumption increasing again.

Autonomy. While higher goals resulted in higher effort, as the author states, ensuring that goals are achievable is crucial in ensuring long-term participation. Therefore, when considering goals in a Rasch-based energy system, long-term effects should be considered, and goals should be effective as well as realistic for a user. And while users might benefit from some guidance in selecting a goal to strengthen their beliefs that a goal would be realistic, user autonomy should also be respected. Eccles and Wigfield (2002) argue, based on several previous studies, that achievements that one feels personally responsible for lead to higher feelings of competence than achievements than those that are reached outside one's own agency. Koestner et al., (2008), also found that goals that were autonomously set, in accordance with one's own motivation and values, led to higher achievement than goals that were externally controlled. Helping a user decide on a reasonable goal, by giving them several goal-setting options to choose from; from easy to more difficult goals, might therefore help to keep goals realistic and within users' possibilities while still providing a certain level of autonomy.

[H1] *Guided goals in a Rasch-based energy recommender result in higher energy savings than no goals.* -- (Refer to the theoretical model in Figure 3 below this section).

While the effectiveness of such a system is important, user satisfaction is equally so; an optimal but unused system will achieve nothing whereas a frequently used, sub-optimal system still has opportunities. In a previous energy-saving recommender study by Knijnenburg et al., (2014), it was found that all eventual effects of system design on performance and choice satisfaction were mediated by system satisfaction. They found that the effectiveness of various preference elicitation methods within such a system (e.g., letting users directly or indirectly indicate their needs), depended on the extent of energy-conservation domain knowledge of the user. The number of selected measures, as well as choice satisfaction, were completely mediated by system satisfaction in three out of four studies and largely mediated in the fourth study.

In our system, it might be that outcomes such as kWh savings and choice satisfaction are partly or completely mediated by system satisfaction, goal support, and experienced choice difficulty. There might also be direct effects of objective system aspects on outcomes.

[RQ2] *To what extent can the effects of system characteristics on choice satisfaction, energy-saving self-efficacy, and kWh savings, be explained by system satisfaction, goal support, and choice difficulty, and how are these effects moderated by personal characteristics such as pro-environmental values and money importance values?*

To understand possible effects of goal setting on user satisfaction within a recommender system, the study by Schäfer and Willemsen (2019) on a nutritional recommender system might be helpful. This study involved Rasch-based personalized goals in a nutrition recommender system. The personalization here was accomplished by bringing certain nutrients into focus that would be reasonable targets for the user. For reachable goals, small improvements in user nutrition behavior were observed after two weeks, especially for those goals that were brought into focus. It was found that an increase in Rasch difficulty level decreased the success rate of items. It was also found that tailored recommendations increased perceived diversity, which had a two-fold effect on effectiveness: a direct negative effect, but an indirect positive effect through increased system satisfaction.

Because goals allow people to see how their chosen measures contribute towards reaching their goal, it is expected they will choose items more deliberately, ultimately being more satisfied with their choices as they have better justifications for their choices. A goal that serves as a focal point might also make it easier to compare measures and therefore reduce could reduce choice difficulty. Together, these effects might lead to higher system satisfaction.

[H2] *Goal setting will improve choice satisfaction.*

[H3] *Goal setting will reduce perceived choice difficulty.*

[H4] *Goal setting will improve satisfaction with the system.*

A system that aids users better in saving energy is expected to be perceived as better supporting users in their energy-saving endeavors. (Assuming these are present based on previously mentioned research by Milieu Centraal & Motivaction, 2023). We will refer to this as 'goal support'.

[H5] *Goal setting will improve perceived goal support.*

Energy-Saving Self-Efficacy. Lastly, we expect an effect of goal setting on energy-saving self-efficacy. Bandura (1997), described self-efficacy as the confidence one has in one's ability to *"organize and execute a given course of action to solve a problem or accomplish a task"*, as cited by Eccles and Wigfield (2002, p. 110). In the context of energy saving, we are interested in the extent to which an energy-saving recommender system helps strengthen users' beliefs in their ability to save energy. This effect might be twofold: It was previously found that the presence of a goal can improve self-efficacy in participants (Bandura and Schunk, 1981); likely because goals can provide one with a way to evaluate one's performance (Elliot et al., 1994). A condition therein is that the goal is indeed reached. This would be a direct effect. On the other hand, increases in system satisfaction and perceived support might indirectly increase self-efficacy; if users feel confident that they can reach certain goals because of the system supporting them. It was previously found that when users are given a goal, their performance more strongly impacts self-efficacy than when there is no goal present (Cervone et al., 1991).

[H6] *Goal setting will improve energy-saving self-efficacy.*

2.8. Signposting

What might also be of interest, is attribute translation of saving goals and saving metrics, also referred to as 'signposting'. With signposting, the same information is presented with certain different units, which are in turn expected to activate certain existing values in users and thereby influence behavior (Ungemach et al., 2018). Signposting is different from the framing interventions discussed before by Starke et al. (2021), and Bams (2018) because it does not rely on valence differences, but purely on a different presentation of information (Ungemach et al., 2018). In the current energy recommender, goals and saving metrics can be presented either as kWh goals, monetary values, or CO2 emission reductions. Energy usage of actions was previously displayed as kWh per year for each item, while investment costs were displayed as monetary costs. This raises the question of whether the way in which these attributes are presented could influence energy savings and user satisfaction.

[RQ3]: What is the effect of signposting on choice satisfaction, self-efficacy, and energy saving in a Rasch-based energy-saving recommender system?

Stadelmann and Schubert (2018) studied the effect of displaying household appliance energy usage as either monetary or kWh costs in an online shop, e.g. the energy usage of fridges and washing machines. They found no difference between these labels: both labels were effective in increasing the sales of energy-efficient appliances. The authors state that this might in part have been due to the usage of a familiar energy label (EU standard) compared to an unfamiliar monetary label. They furthermore found that while the purchased appliances were more efficient, they were also larger and thus eventually would consume the same amount of energy. However, they did not distinguish between different kinds of customers who might be affected differently by different labels.

Ungemach et al. (2018) studied signposting in a car comparison task, depicting fuel consumption as either greenhouse gas emissions or annual fuel cost. This had a different effect on different users, depending on their environmental attitudes as measured by the NEP scale. Users with stronger pro-environmental attitudes opted for more fuel-efficient cars when they were presented with greenhouse gas ratings, rather than monetary ratings. The reverse was true for those with lower NEP scores. Ungemach et al. (2018) hypothesized that signposting works by activation of certain values, without the need for manipulating users, as they are presented with the same information which is merely translated differently. This was supported by the observation that the signposting effect diminished when environmental values were activated in different ways, e.g., through first letting users fill out the NEP scale, or by educating them on the topic before presenting them with the task. In those cases, people with higher NEP scores tended to choose the energy-efficient cars, regardless of how attributes were presented. Nonetheless, given that our energy recommender is quite ambiguous with regards to environmental vs. monetary values, we do not expect strong activation of these values through other means than through signposting. Therefore, signposting of saving metrics in our system might have similar effects as Ungemach et al. (2018) found.

Other Effects. In the below section, we will introduce several more hypotheses. While we have some general expectations concerning the direction and type of effects we expect to find (from objective system aspects, to subjective system aspects, to experience and interaction outcomes, in line with the user-centric evaluation framework by Knijnenburg and Willemsen, 2015), we are unsure of the exact individual effects, and these can in large be perceived to be in exploratory nature (Except for hypothesis 7 and 8, which we have stronger expectations for).

We expect that signposts that are in line with user values, on a pro-environmental dimension and on an importance of money dimension, will result in higher energy savings; thus, we expect that monetary (Euro) signposts will result in higher savings for those with stronger financial values as measured by the importance of money (IMS) scale, and we expect that the CO2 signpost will result in higher savings for those with stronger pro-environmental values, as measured by the New Environmental Paradigm (NEP scale). We have no expectations for the kWh signpost in terms of value activation. Therefore, we use this signpost as a baseline with which we compare the CO2 and Euro signposts. This would also enable us to draw conclusions as to which signpost is beneficial for which type of user.

[H7] *CO2 signposting, as compared to kWh signposting, will result in higher savings for increasing strength of people's financial values.*

[H8] *Monetary signposting, as compared to kWh signposting will result in higher savings for increasing strength of people's pro-environmental values.*

Additionally, as users have previously been shown to prefer tailored systems (Starke et al., 2017, Starke et al., 2021), we expect that signposting in accordance with user values, will increase user satisfaction. In the past, the option to sort on saving metrics, or to display items as either kWh or Euro savings, was used by less than half of the participants. (Knijnenburg, 2009). However, there might have been a group for which the interface was already optimal in terms of signposting, thus leaving only a smaller group that seems indifferent to the signpost used. As with the goal-setting hypotheses, we expect that the effects of signposting will result in reduced choice difficulty, higher choice satisfaction and system satisfaction, increased goal support, and stronger environmental self-efficacy when the signposts are in line with user values. Because CO2 information might be more relevant to those with stronger pro-environmental values, it could aid their decision-making better, as compared to those with weaker pro-environmental values. This will then also increase perceived system satisfaction, choice satisfaction, goal support, and self-efficacy.

[H9, H10, H11, H12, H13] *CO2 signposting, as compared to kWh signposting, will result in **[H9]** reduced choice difficulty, **[H10]** higher choice satisfaction, **[H11]** increased system satisfaction, [**H12]** increased goal support and **[H13]** stronger environmental self-efficacy with increasing strength of pro-environmental values.*

The same argument could be made for financial values (importance of money) and their effect on user satisfaction across several constructs.

[H14, H15, H16, H17, H18] *Monetary (Euro) signposting, as compared to kWh signposting, will result in* ***[H14]*** *reduced choice difficulty,* ***[H15]*** *higher choice satisfaction* ***[H16]*** *increased system satisfaction,* ***[H17],*** *increased goal support and* ***[H18]*** *stronger environmental self-efficacy with increasing strength of financial values.*

We must note that a similar study on attribute translation and goal setting was conducted previously by Brandsma and Blasch in 2019, in which they found no effect of attribute translation on overall willingness to conserve energy. However, they did find differences between users with several different value orientations for their willingness to save energy. Those with stronger biospheric values were more motivated to save energy, while those with more egoistic values were less willing to reduce their energy usage, except when savings were displayed as monetary savings.

However, there are various reasons why we think that this study could be improved upon, and why our context sufficiently differs to justify our current study. The study by Brandsma & Blasch (2019) only considered a single energy-saving measure: turning off stand-by appliances on a daily basis. This measure was given a low difficulty rating in the study by Bams; several such actions (e.g., turning off the computer completely) were considered to be quite easy (around difficulty level 0). Therefore, there is a decent chance that this action is already being performed by around 50% of users, making it more difficult to observe an effect. Furthermore, participants were only asked to imagine setting themselves a goal, and then say if they were willing to perform this action. The environmental attribute translation was furthermore displayed as the equivalent of driving a certain distance by car. Thus, if someone is unaware of how much greenhouse gas a car emits over that distance, they are none the wiser in terms of environmental impact.

In addition to the effects of signposts and goal conditions on outcome variables and subjective experiences of users, we expect that the subjective experiences will act as mediators between the objective system aspects and the user experience and interaction outcomes. To be precise, we expect the following relationships:

[H19, H20, H21, H22] *Reduced choice difficulty will lead to* ***[H19]*** *an increase in energy-saving self-efficacy,* ***[H20]*** *an increase in choice satisfaction,* ***[H21]*** *increase in kWh savings, and* ***[H22]*** *an increase in system satisfaction.*

[H23, H24, H25] *Increased system satisfaction will lead to* ***[H23]*** *an increase in energy-saving self-efficacy,* ***[H24]*** *an increase in choice satisfaction, and* ***[H25]*** *an increase in savings.*

[H26, H27, H28, H29] *Increased goal support will lead to* ***[H26]*** *an increase in energy-saving self-efficacy,* ***[H27]*** *an increase in choice satisfaction,* ***[H28]*** *an increase in savings, and* ***[H29]*** *an increase in system satisfaction.*

Additionally, similarly to what was seen in the research from Knijnenburg and Willemsen (2014) on preference elicitation in energy-saving recommender systems, we also expect that increases in (chosen) savings will lead to increases in choice satisfaction. We also expect that increases in choice satisfaction will translate into stronger feelings of energy-saving self-efficacy.

[H30] *Increased savings will lead to increased choice satisfaction.*

[H31] *Increased choice satisfaction will lead to increased energy-saving self-efficacy.*

Perceived Feasibility. Another factor that might be influenced by this system, is perceived feasibility. Starke et al. (2017) found that easier measures were perceived as more feasible than more difficult measures. They also found that increased perceived feasibility led to increased choice satisfaction.

[Exploratory RQ] *What is the effect of goal setting on perceived feasibility in a Rasch-based energy recommender system?*

For the current study, the effect of goal setting on perceived feasibility might be twofold, and hypothesizing a direction is therefore complex; Chosen items might be perceived as more feasible when a user has a stronger belief in their ability to save energy (resulting from a goal), but at the same time, a user might pick more difficult items, thus decreasing the objective feasibility of items. As an exploratory hypothesis, we expect that these will cancel each other out.

[Exploratory Hypothesis 1] *Goal setting will result in the same items being perceived as more feasible.*

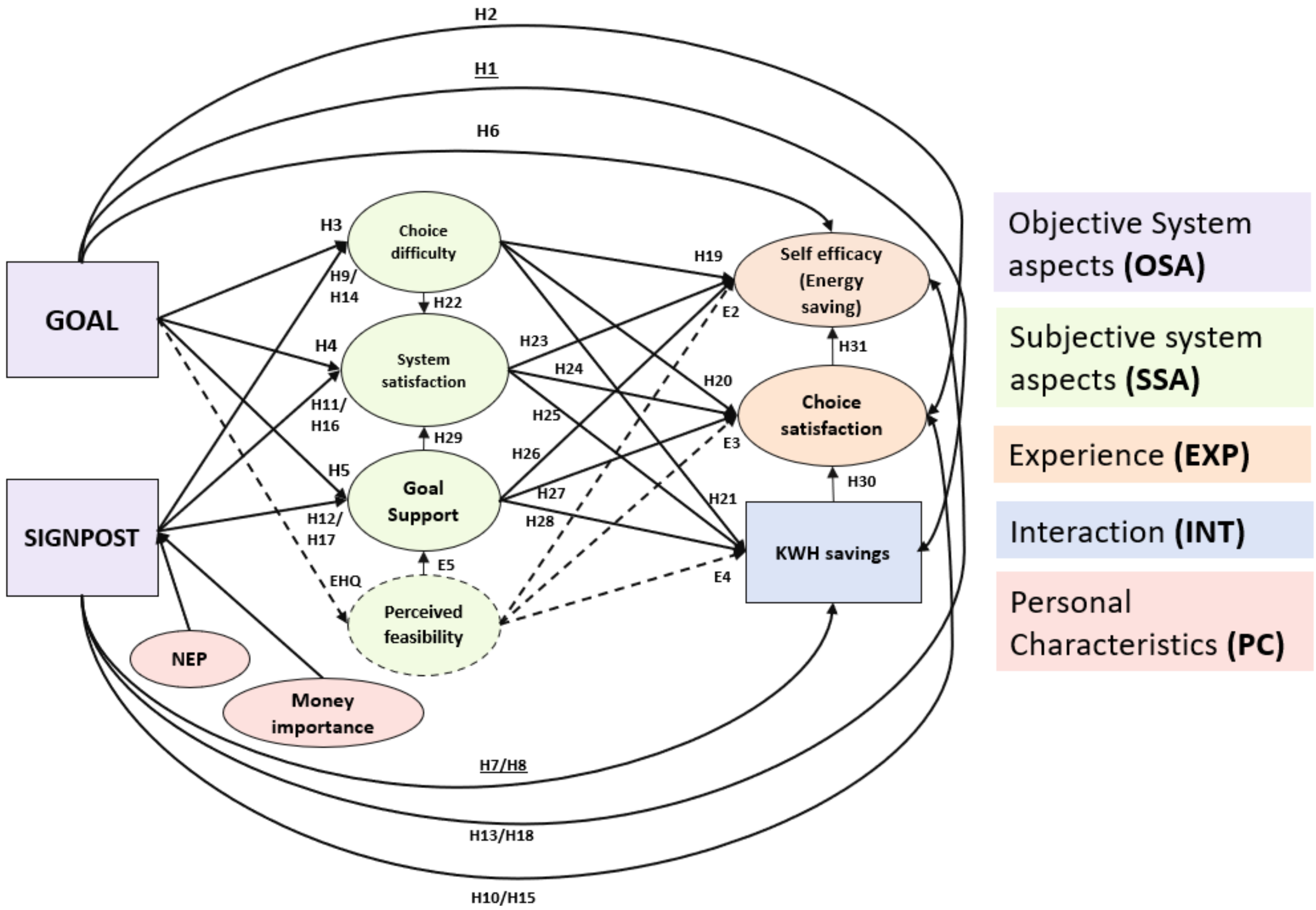
We furthermore have several expectations about the effects of perceived feasibility on system outcomes, in line with the effects of other subjective system aspects.

[E2, E3, E4, E5] *We expect positive effects of perceived feasibility on* ***[E2]*** *self-efficacy,* ***[E3]*** *choice satisfaction,* ***[E4]*** *savings, and* ***[E5***]*goal support.*

All hypotheses so far can be captured with the below model (figure 3), with the dashed node and lines depicting the exploratory hypotheses. The observed variables are depicted as squares, and the latent variables as ellipses. The model generally follows the user-centric evaluation framework by Knijnenburg and Willemsen (2015), which describes the relationship between objective system aspects (on the left), and interaction outcomes and user experiences (on the right), through various subjective system aspects (in the middle). These effects might be moderated by situational characteristics and personal characteristics, the latter of which would be environmental and financial values in our model. The main difference between the original framework and our current model, is that we consider self-efficacy to be an outcome measure, whereas usually it would be considered a personal characteristic that influences user experiences.

Figure 3

Theoretical model with all hypotheses



3. Methods

3.1. Experimental Design

The study employed a 3*2 between-subject design (see Table 1). Participants were divided into a *goal* and *no goal* condition, and into three signpost conditions. In these signpost conditions, attributes and goals were translated as either kWh, monetary (Euro), or kg CO2 savings metrics.

Table 1

Study design

Signpost	Goal	No Goal
kWh	kWh + goal	kWh
Monetary	Monetary + goal	Monetary
Kg CO2	CO2 + goal	Co2

3.2. Participants

In total, 212 people participated in the initial study, of which 94 were recruited through the Prolific participant database and 117 through the Archie participant database from Eindhoven University of Technology. Additionally, one person participated as an external participant, after being invited by someone else. We removed one participant for not currently residing in the Netherlands, three were removed because they showed very little variation in scale answers, i.e., only clicked the same (non-zero) answer for all items in the system satisfaction scale, while some of these items were phrased as opposing statements. One observation was removed because a person participated twice. Additionally, we removed five additional participants for selecting more than 16 energy-saving measures they intended to take (further than 2 *SD* from the mean), as we considered this to be an unrealistic amount, and thus we regarded these as outliers. After this, 202 participants remained, of whom 170 participated in the follow-up study. For the follow-up, we also removed 10 participants for various reasons (which will be explained in the corresponding results section). In Table 2, you can see the number of participants per condition, for the initial study and the follow-up study.

Table 2

Study design and participants per condition, per study (after outlier removal)

	Initial study			Follow-Up study		
Signpost:	Goal	No goal	Total	goal	No goal	Total
kWh	32	35	67	25	28	53
Monetary	31	41	72	21	34	55
Kg CO2	34	29	63	29	23	52
Total	97	105	202	75	85	160

3.1.1. Description of the Sample

The remaining sample of 202 participants consisted of 86 males, 112 females, 1 other, and 3 undefined, with a mean age of 30.5 (*SD* = 14.7) and a median age of 25. 75% of participants were younger than 32, not differing much between Prolific and Archie databases. 60% of participants obtained university-level education, 22% obtained education from a university of applied sciences, 10% obtained vocational education, 7% obtained high school education, and 2% obtained no education or did not wish to disclose this. Participants were, other than the one removed, exclusively currently living in the Netherlands, as this was a selection criterium on both Prolific and Archie.

Of the participants, 23% were homeowners, 74% rented living space, and 3% neither owned nor rented accommodation, e.g., lived with parents or elsewhere. 22% lived in a (partly) freestanding house, 26% in a terraced house, 33% in an apartment, and 17% in a room. Most participants did not know or did not want to disclose the energy label of their house (56%), and otherwise, 10% indicated energy level A, 9% indicated B, 12% indicated C, 4% indicated D and the remaining 9% indicated energy label D or below. 53% lived in a dwelling built before or during 1991 whereas 24% lived in a dwelling built during or after 1992, and the remaining 23% did not answer this question.

All in all, 57 out of 202 participants rented an apartment, 22 owned a (partly) freestanding house, and 15 owned a terraced house. Of the people who owned a (partly) freestanding or terraced house, 24% had energy label A, 11% indicated energy label B, and 30% indicated energy label C. 35% of participants chose English as the experiment language and 65% chose Dutch. Approximately 23% of participants participated on a phone or tablet (where screen height was larger than screen width).

3.1.2. Statistical Analyses and Sample Size Calculations

To determine the difference between the goal and no goal condition on energy savings, a t-test was used. Additionally, goal conditions were considered in a regression together with signposts and interaction effects. To determine the sample size, we considered two studies on goal setting in the energy-saving context. Abrahamse et al., (2007) found a large effect (Cohen's d = 0.85) for direct energy consumption when participants were instructed to save 5% energy over a period of 5 months, However, Harding and Hsiaw (2015) found only a small effect in the study where participants selected energy saving measures themselves, without guidance on what would be realistic plans. Therefore, we expected a moderate effect of d=.5, as participants would have some guidance when setting their goals on what would be an accessible, moderate, and challenging goal. This would have required 140 participants according to G-Power software.

To measure the direct effect on energy savings between signposts, (averaged over goal conditions), an analysis of variance (ANOVA) over 3 groups was used, as well as a regression with interaction effects. While Ungemach et al. (2018) found an effect size of .15 for their study on signposting, we aimed to be able to show slightly larger effects. This is because the study by Ungemach on signposting was qualitatively very different, concerning a choice task between two cars, which conveyed both the cars' prices and the fuel consumption signposts. Due to more distinct experimental conditions that primarily show the signpost in our study, as well as the very different nature of our task, we expected a larger effect. Furthermore, a sample size of 500+, which would be needed for an effect size of .15, would be very expensive, difficult to gather, and such a small effect might not have practical relevance: Signposting effectiveness depends on value orientations, which would first have to be determined before signposting becomes effective. A small effect might not be worth the effort of measuring these values. We powered our study such that we could show medium effect sizes at best. This would require 207 participants according to the G-power software.

Lastly, to test the theoretical model depicted in Figure 3, a structural equation model (SEM) was used. To determine an appropriate sample size, the SEM sample size calculator by Daniel Soper (2020) was used. Based on a medium effect size, around 200 participants would have been needed for 30 observed variables with 5 latent variables.

Given the largest required sample size of 207 to observe the effect between signposts in the ANOVA, we aimed for a total sample size of 210.

3.2. Stimulus Material

Measures were obtained from a previous study by Bams (2018) and largely overlap with the measures used in previous studies described in the Ph.D. thesis of Starke (2019). The list consisted of 135 energy-saving recommendations with varying levels of difficulty. These difficulties were determined in previous studies that asked a large panel of participants which measures they performed: A measure that was performed by fewer participants, resulted in a higher difficulty score in this Rasch-based system. Then, someone who performs few measures, and thus obtains a lower ability rating, will be presented with items that also have a lower difficulty rating. This will be further explained in the procedure section. For the full list of measures, saving indices, difficulties, and whether they were counted as a gas or electricity measure, refer to Appendix A.

3.2.1. Saving Aid Application

For this study, the 'Saving Aid' web app was launched in which people could obtain personalized energy advice in a Rasch-based energy-saving recommender system. The workings and design of the recommender interface were in part based on reports of earlier studies, e.g. Starke (2021) and Bams (2018), but otherwise, the website was built entirely anew for this study (in React.js and PHP/SQL), and no code was re-used from these earlier studies as this was no longer available. For additional explanations of the workings of this app (Front-end and back-end), and screenshots of the overall recommender interface in desktop and mobile modus, refer to Appendix B.

3.3. Procedure

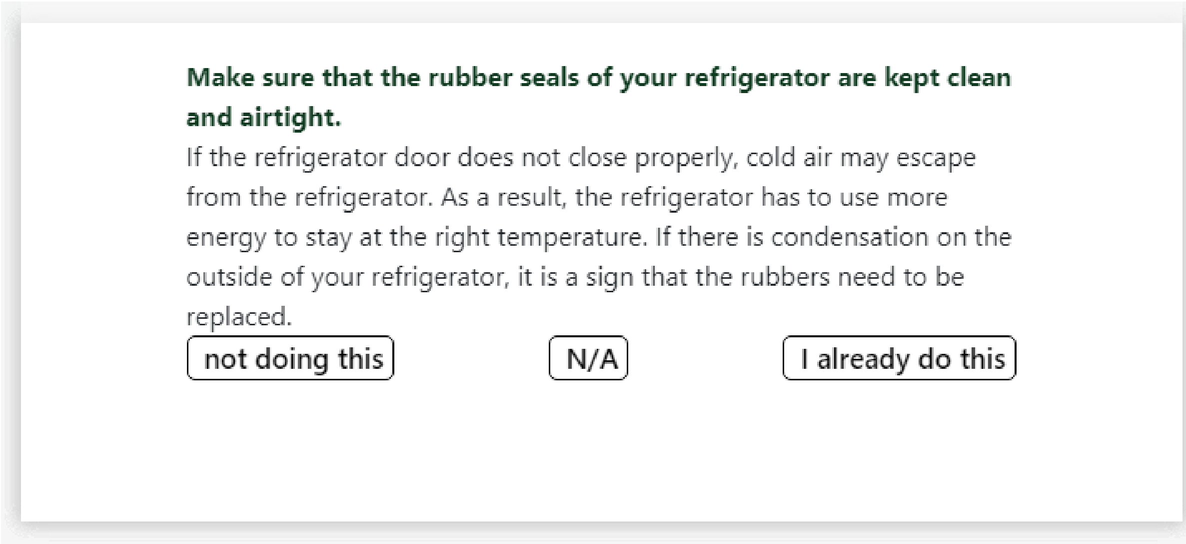
Participants completed the entire study within the saving aid interface. This took 15 minutes for the initial recommendations and a maximum of around five minutes for the one-month follow-up study. The initial study was run from May 11th, 2023, to May 31st, 2023, and the follow-up study was run from June 15th, 2023, to June 23rd, 2023. For a visual walkthrough of the experiment with screenshots of each screen, refer to Appendix C.

The Saving Aid web app was otherwise structured like a regular online experiment, which started with a screen where participants could choose their language, either Dutch or English. (In order to see possible differences between language conditions afterwards, we did not allow participants to change this during the experiment.) Translations were manually checked and implemented to ensure consistency between the two language conditions.

Then, participants were presented with a consent form, followed by a survey on current energy-saving behaviors. For this, participants were asked which measures they currently performed, out of 19 semi-randomly chosen measures. To obtain these lists of 19 measures, all 135 measures were sorted on difficulty and divided into 19 batches. From each batch, one measure was randomly presented to the participant (and these were shown in random order). Participants were asked to indicate if they performed these measures or not, or if the measures did not apply to their situation. An example item is displayed below in Figure 4.

Figure 4

Sample of a current item



Ability Calculation. Based on these responses, participants were assigned an ability score, corresponding to the average difficulty of the Nth batch of items, which was based on the total number N of measures that a participant currently performed. A correction was made in case participants selected N/A (Not applicable). In that case, the Nth batch was calculated as depicted by formula 2 below.

$$\textbf{N/A-corrected Nth batch} = (\text{number of items} / (19 - \text{number of non-applicable items}))* 19 \quad (2)$$

Resultingly, someone who only selects N/A and "Yes", i.e., who does all that is possible in their situation, will end up with the highest possible ability score.

Goal Setting. After this, half of the participants were presented with the task of choosing an energy-saving goal, in which they could choose amongst three goals; an 'accessible' goal of 600 kWh, a 'moderate' goal of 1200 kWh, and a 'challenging' goal of 1800 kWh. An example is depicted in Figure 5. The saving attributes of these goals were translated based on signposting conditions: ~1/3 of the participants in the goal-setting condition saw a monetary savings goal, ~1/3 saw a CO_2 reduction goal, and ~1/3 saw a kWh savings goal by random assignment of the system. The monetary goals were translated to €180,-, €360,-, and €540,- respectively, and the CO_2 reduction goals were translated as 150, 300, and 450 kg CO_2 respectively. It was indicated for the three goals, that these would *"take relatively little time and cost relatively little"*, *"take a moderate amount of time and cost a moderate amount"* and *"take relatively much time and cost relatively much"* to achieve. The height of these goals was based on an overall 1200 kWh average savings in the previous study by Bams (2018), and a median of 660 kWh for those who chose at least some amount of savings in that study (obtained from previous data, not from the publication). We only considered the group who chose at least some amount of savings, because we expected the portion of people who would not choose any savings, to be smaller than in 2018 (where at least 25% of participants chose no savings at all), given the research by Milieu Centraal (2023), and recent developments in energy prices.

Figure 5

Example of goal

Choosing New Measures. Next, all participants were presented with an instruction screen that showed and described in 5 steps what was expected of participants as shown in Figure 6. They were asked to pick items they intended to perform in the future, to remove items they already performed (after which new ones would appear at the bottom of the list), and were shown how to obtain more information on measures (by hovering over the images or clicking the 'more info' buttons).

Figure 6

Instructions for participants

Recommendation Algorithm. After this, participants were presented with a tailored list of energy-saving recommendations. These were the items with difficulty scores closest to the participant's ability, sorted on the absolute distance of ability score and item difficulty. Therefore, going down the list, both increasingly more difficult and increasingly easier items appeared, compared to those at the top of the list. The list was initially 20 measures long, with a maximum of 10 replacements shown at the bottom of the list for each removed item. Example measures are displayed below in Figure 7.

Figure 7

Example items from the recommender list.

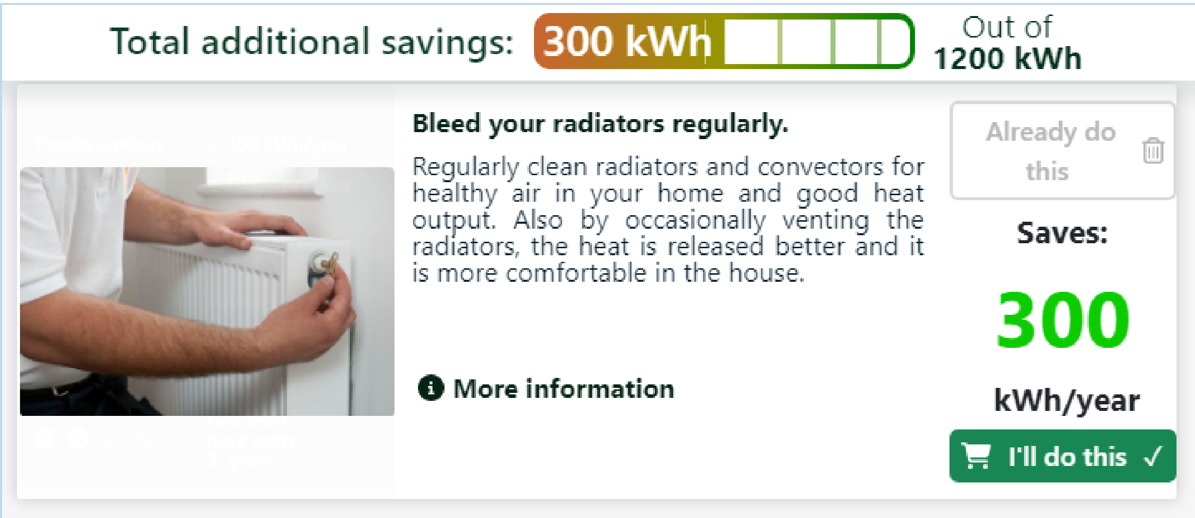
The bottom measure shows the information container that appears on clicking 'more information' or hovering over the image.

Signposts. The measures were accompanied by highlighted saving metrics as seen on the right in Figure 7, which were translated as either a kWh, monetary, or CO2 metric depending on the signpost condition the participant was in. CO2 metrics were calculated as 0.25 kg CO2 per kWh, and the monetary metrics as €1.45 per m^3 gas (With an average of 0.102 m^3 per kWh, resulting in ~€0.15 / kWh) for gas-based measures, and €0.40 per kWh for electricity-based measures in line with current price ceilings in the Netherlands. Although this translation difference between the two types of measures is not entirely in line with the signposting theory (which states translations should present the exact same information differently), this was deemed necessary due to the large cost differences in kWh and m^3 gas prices: This ensured that participants were not grossly misled about potential monetary savings.

Goals. Additionally, the participants who chose a goal, saw this goal displayed at the top of the page as in Figure 8, in the same unit as the signpost. The monetary goal was translated as €0.30 per kWh, close to the weighted average of the kWh and Gas prices. Due to the different costs per kWh for gas- and electricity-oriented measures, it could happen that with the same items, the monetary goal would have been achieved, whereas the CO2 or kWh goal would not. However, with some testing, the €0.30 per kWh seemed to align quite well with the kWh / CO2 goals, and the progress bar acted similarly for all three signpost conditions.

Figure 8

Goal setting condition (With kWh signpost)



Participants who were not in the goal condition saw their total savings at the top of the page as depicted in Figure 9. This was done because the mere presence of the total could influence savings, and we were interested in the effect of goal setting as an isolated variable. A previous study showed that merely displaying total water usage influenced behavior (Tiefenbeck et al., 2016). Thus, displaying the total in both conditions allowed us to see the effect of goal setting in isolation.

Figure 9

No goal condition (with Euro signpost)

Questionnaires. After having chosen measures from the list, participants were asked about their experience with the system as well as their personal values. First, they were asked about the extent to which they experienced choice difficulty, how feasible they perceived the items to be, to what extent they were satisfied with their choices, the system in general, and to what extent they believed the system would support them in saving energy in the future (goal support). Then, three questionnaires were shown about their beliefs in their ability to save energy, as well as the New Environmental Paradigm (NEP-scale) on pro-environmental values, and the Money Importance Scale (Refer to the measures section below).

Demographics. Subsequently, participants were asked various demographic questions about their age, gender, education level, and current country of residence, as well as questions about their living conditions: Whether they owned a house or not, in what type of house they lived, and what the energy label and building year of the house was.

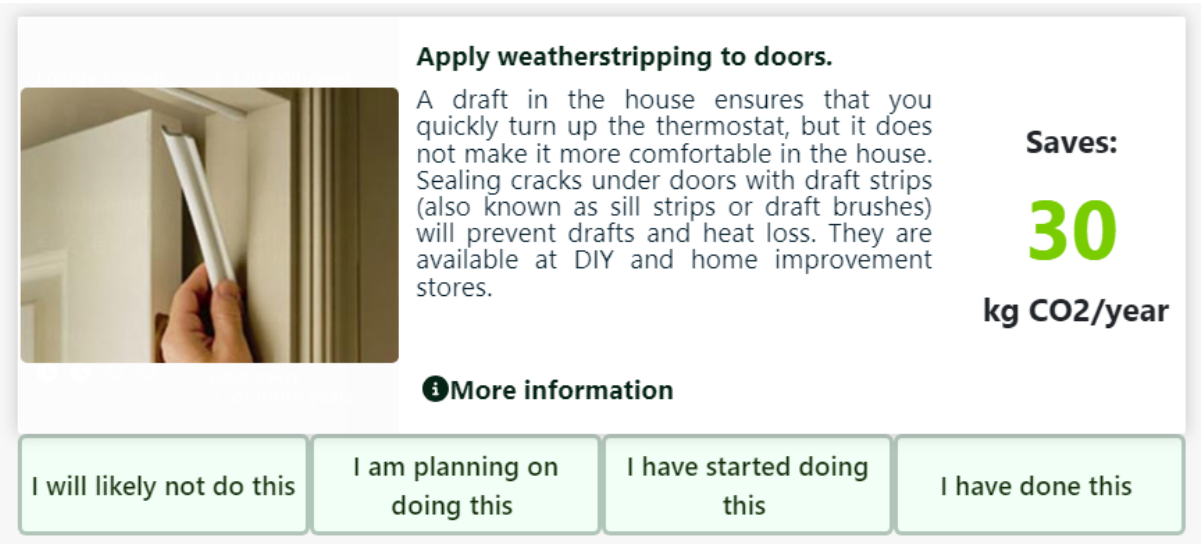
Review and Closing Screen. After having answered all questions, participants were shown their chosen measures and given the option to send themselves the link to view them in the future. Lastly, participants were shown a thank-you screen in which they could invite someone else, sign up for updates about the research, or leave additional feedback about the system. Prolific participants were shown a link to return to Prolific.

Review Email. Given that the review link was not used by many participants, and we noticed some participants returned to the experiment at a later time (possibly in search of their recommendations/chosen items), we decided to send an email approximately one week after the experiment with a personalized link to give everyone access to their chosen measures. These measures had the same signposts as in the initial study, but not the goals or other interactive functionality.

Follow-up. Around four weeks after the initial study, participants were invited to a follow-up experiment in which they were asked to what extent they had really performed their chosen measures. They could choose amongst *'I will likely not do this'*, *'I am planning on doing this'*, *'I have started doing this'*, and *'I have done this'*, as seen in Figure 10.

Figure 10

Follow-up sample



3.4. Measures

In total, eight different scales were used; one for each latent construct in the theoretical model in Figure 3 (depicted with ellipses). For all scales, a 7-point Likert scale (from strongly disagree to strongly agree) was used, except for the money importance and NEP scales, which used 5-point Likert scales. Refer to Table 3 below for all items and factor loadings in the initial and follow-up study, and to Appendix D for a full overview of item sources.

Choice Difficulty. To measure choice difficulty, several items from Willemsen et al. (2016) were used. This scale originally had an Alpha of .73 and AVE (Average Variance Explained) of .56. Some example items included: *"It was easy to choose between energy saving measures"*, *"I changed my mind several times while choosing energy saving measures "* and *"The task of choosing energy saving measures was overwhelming".*

Perceived Feasibility. Perceived feasibility was measured using three items from the second study by Starke et al. (2017), which originally had an Alpha of .83**.** These items were: *"I think it would take me little effort to perform the chosen measures."*, *"I do not have the possibility to perform the chosen measures."* And *"I think the chosen measures are easy to apply in my home environment."*

Choice Satisfaction. The choice satisfaction scale used items from Willemsen et al., 2016, which had an Alpha of .93 and AVE of .85, and statements from Knijnenburg et al., 2014, which had an AVE of .54. Sample items include: *"I am satisfied with the measures I chose"* and *"I think I would enjoy performing the chosen energy saving measures".*

System Satisfaction Scale. The system satisfaction scale was composed of three different scales: The system satisfaction scale from Knijnenburg et al. (2014), the system effectiveness scale from Knijnenburg et al. (2012), and the perceived support scale from Starke et al. (2017), which had an alpha between .81 for the first study and .92 for the second study. Items included *"I make better choices using the Saving aid."*, "*The Saving aid is helpful to find appropriate measures."* And *"I would recommend the Saving aid to others".*

Goal Support. The goal support scale consisted largely of original items, that purposefully omitted the word *'goal'*. This was done to avoid confusion amongst those who were in the goal-setting condition. Instead, it assumes some degree of motivation towards conserving energy on the part of the participant and asks directly how the participant thinks the Saving Aid helps in achieving that pursuit. Example items include *"The Saving aid makes saving energy easier", "The Saving aid motivates me to save more energy.",* and, *"I think I will save more energy in the coming year thanks to the Saving Aid."* Two items from Knijnenburg et al. (2012) and Knijnenburg et al., (2014) were added to the goal support scale because they seemed closer related to goal support than to the system satisfaction scale. One of these statements was *"The Saving aid makes me more aware of my options for saving energy".*

Self-Efficacy Scale Regarding Energy Saving. This scale was obtained from Lee and Tanusia (2016), who adapted this from the 'general self-efficacy scale' by Chen et al. (2001). Items included *"I will be able to achieve most of the goals that I have set for myself concerning energy conservation."* and *"When facing difficult decisions on energy conservation, I am certain that I will accomplish them.".*

Pro-Environmental Values. For measuring pro-environmental values, the widely used revised NEP-scale was used (New environmental paradigm), by Dunlap et al. (2000). This scale has an Alpha of .83. Example items of this scale include *"We are approaching the limit of the number of people the earth can support"* and *"The balance of nature is strong enough to cope with the impacts of modern industrial nations".*

Monetary Values. To measure monetary values, the *"importance of money scale"* was used (IMS), as devised by Franzen and Mader (2022). This scale has an alpha of .82. It is the first paper to separate the concept of materialism and the extent to which one values money itself. The paper was only cited one time, but given it was a very recent study and the good alpha score, we decided to use this scale. Items included *"Financial security is important for my well-being."* and *"One can only have a decent life with a lot of money.".*

3.4.1. Factor Analysis.

To verify the proposed constructs of our theoretical model, we performed an exploratory factor analysis (EFA) in the mPlus software package (Muthén & Muthén, 2023). Items with low factor loadings (CDIF2) or high cross-loadings (Several goal support, choice satisfaction, and system satisfaction questions) were removed from the subsequent CFA and SEM analyses. From the EFA, we determined that goal support questions 3, 4, and 5 with system satisfaction question 2, measured a single construct. To simplify the analysis somewhat, we used the unweighted computed scores for the money importance scale (IMS) and new environmental paradigm (NEP), rather than using the individual questions for the CFA/SEM. This was done because these were existing and validated scales, and their statements were not directly related to the system itself. We therefore did not expect that these scales would render vastly different outcomes in our study. The confirmatory factor analysis of the remaining scales can be seen in the methods section, Table 3. The factor loadings from Stata for the NEP and IMS scales can be found in Table 4. All scales in Table 3, have an average variance explained rating (AVE) of at least 0.45, not considering the greyed-out items that were omitted from the analysis. Perceived feasibility, a construct that was intended for the exploratory hypothesis, was not part of the final model due to a high correlation with choice difficulty.

Table 3

Confirmatory Factor analysis (mPlus results) with Alpha and AVE for study (1) and (2)

Construct	Name	Statement	Study 1		Study 2	
			Factor loading	R^2	Factor loading	R^2
Choice difficulty	**CDIF1**	It was easy to choose between energy saving measures	**-.73**	.56	**-.83**	.71
	CDIF2	I changed my mind several times while choosing energy saving measures				
	CDIF3	The task of choosing energy saving measures was overwhelming	**.53**	.30	**.50**	.28
Alpha:**.66**(1)/**.67**(2) AVE:**.45**(1)/**.47**(2)	**CDIF4**	Comparing the energy saving measures took a lot of effort	**.73**	.50	**.63**	.43
Choice satisfaction	**Chsat1**	I am satisfied with the measures I chose	**.49**	.66	**.51**	.78
	Chsat2	I think I would enjoy performing the chosen energy saving measures	**.41**	.47	**.39**	.46
	Chsat3	I would recommend the chosen measures to others	**.47**	.59	**.49**	.72

	Chsat4	I will implement all the measures I have chosen				
Alpha:**.73**(1)/**.76**(2) AVE:**.66**(1)/**.65**(2)	Chsat5	I think I chose the best energy saving measures from the list				
Goal support	SYSSAT1	I make better choices using the Saving aid.				
	SYSSAT2	The Saving aid is helpful to find appropriate measures.	**0.74**	.62	**.70**	.51
	SYSSAT3	I would recommend the Saving aid to others				
	SYSSAT4	I would use the Saving aid more often if possible				
	SYSSAT5	The Saving aid was useless				
	GSUP1	The Saving aid makes saving energy easier				
	GSUP2	The Saving aid motivates me to save more energy.				
	GSUP3	I think I will save more energy in the coming year thanks to the Saving aid				
	GSUP4	The Saving aid gives me more insight into the energy consumption of devices and systems in my home	**.77**	.68	**.78**	.63
	GSUP5	The Saving aid made me more energy-conscious	**.77**	.69	**.78**	.63
Alpha:.**86**(1)/**.82**(2) AVE:**.68**(1)/**.62**(2)	**GSUP6**	The Saving aid makes me more aware of my options for saving energy	**.78**	.71	**.83**	.72
Environmental Self efficacy	**SEF1**	I will be able to achieve most of the goals that I have set for myself concerning energy conservation.	**.58**	.54	**.57**	.46
	SEF2	When facing difficult decisions on energy conservation, I am certain that I will accomplish them.	**.53**	.46	**.59**	.48

	SEF3	In general, I think I can obtain energy conservation outcomes that are important to me.	**.66**	.69	**.71**	.70
	SEF4	I believe I can succeed at most any energy conservation endeavour to which I set my mind.	**.61**	.60	**.65**	.59
	SEF5	I am confident that I can perform effectively on many different tasks relating to energy conservation.	**.72**	.82	**.73**	.74
Alpha:**.87**(1)/.**85**(2) AVE:.**60**(1)/**.57**(2)	**SEF6**	Compared to other people, I can do most energy conservation task very well	**.52**	.44	**.58**	.47

Table 4

Factor loadings of NEP and IMS scales and alpha scores for both studies (From STATA)

Construct	**Name**	**Statement**	***Study 1***	***Study 2***
New Environmental Paradigm	**NEP1**	We are approaching the limit of the number of people the earth can support	**.39**	**.39**
(Used computed and unweighted score in SEM, not separate questions)	**NEP2**	Humans have the right to modify the natural environment to suit their needs	**-.36**	**-.37**
	NEP3	When humans interfere with nature it often produces disastrous consequences	**.45**	**.40**
	NEP4	Human ingenuity will ensure that we do NOT make the earth unliveable	**-.40**	**-.45**
	NEP5	Humans are severely abusing the environment	**.64**	**.63**
	NEP6	The earth has plenty of natural resources if we just learn how to develop them	**-.37**	**-.40**

	NEP7	Plants and animals have as much right as humans to exist	**.35**	**.30**
	NEP8	The balance of nature is strong enough to cope with the impacts of modern industrial nations	**-.58**	**-.62**
	NEP9	Despite our special abilities humans are still subject to the laws of nature	**.45**	**.38**
	NEP10	The so-called “ecological crisis” facing humankind has been greatly exaggerated	**-.63**	**-.67**
	NEP11	The earth is like a spaceship with very limited room and resources	**.53**	**.50**
	NEP12	Humans were meant to rule over the rest of nature	**-.54**	**-.52**
	NEP13	The balance of nature is very delicate and easily upset	**.47**	**.39**
	NEP14	Humans will eventually learn enough about how nature works to be able to control it	**-.50**	**-.55**
Alpha: **.81**(1&2)	**NEP15**	If things continue on their present course, we will soon experience a major ecological catastrophe	**.70**	**.69**
Importance of money scale	**IMS1**	Material wealth is important for me.	**.73**	**0.72**
(Used computed score in SEM, not separate questions)	**IMS2**	Money is important for me.	**.78**	**0.80**
	IMS3	Money makes me happy.	**.62**	**0.63**
	IMS4	Financial security is important for my well-being.	**.31**	**0.32**
	IMS6	One can only have a decent life with a lot of money.	**.48**	**0.45**
	IMS7	I enjoy material things.	**.54**	**0.52**
Alpha: **.77**(1&2)	**IMS8**	To make more money, I would work more immediately.	**.50**	**0.54**

4. Results

Below we give an overview of the results obtained in this study, starting with the initial study where people chose items, and concluding with the follow-up study where people reported on which actions they had actually performed after four weeks.

4.1. Initial Study.

Of the 248 people who started the experiment, 209 people completed the initial experiment and an additional three made it to the page where they could review their actions, thus also completing the majority of the experiment. As explained in the methods section, from these 212 (partial) completions, ten were removed for various reasons.

4.1.1. Descriptive Statistics

On average, participants indicated for 8.5 out of 19 measures that they currently performed them, with an *SD* of 2.8. In the screen where participants were asked to choose new measures, participants picked on average 7.5 measures with an *SD* of 3.8, and they removed 10.3 measures with an *SD* of 6.5, indicating that they already performed them. 33 people indicated for every item that they either already performed the measure, or that they were planning on performing the measure in the future; thus, not leaving any items unclicked. Participants looked at the information overlays of 1.9 measures on average (either by clicking on the button or hovering over the image for at least 1.5 seconds), with an *SD* of 3.7. Participants only deselected 0.4 items on average, with an *SD* of 0.86, indicating they did not alter their decisions very often.

In general, participants seemed to be quite satisfied with the system and their chosen measures. On a scale of -3 to 3, the average response for '*I am satisfied with the measures I chose*' was 1.64, with an *SD* of 0.94. Participants responded with -1.95 (*SD* = 1.19) to the statement "*the saving aid was useless*" (Although this statement was not included in the final goal support/system satisfaction factor), and with 1.594 (*SD* = 1.02) to the statement that "*The Saving Aid is helpful to find appropriate measures.*" (on the same scales), indicating that participants were overall content with the system. For a full overview of questions and average answers (with *SD*s, minimums, and maximums) refer to Appendix E.

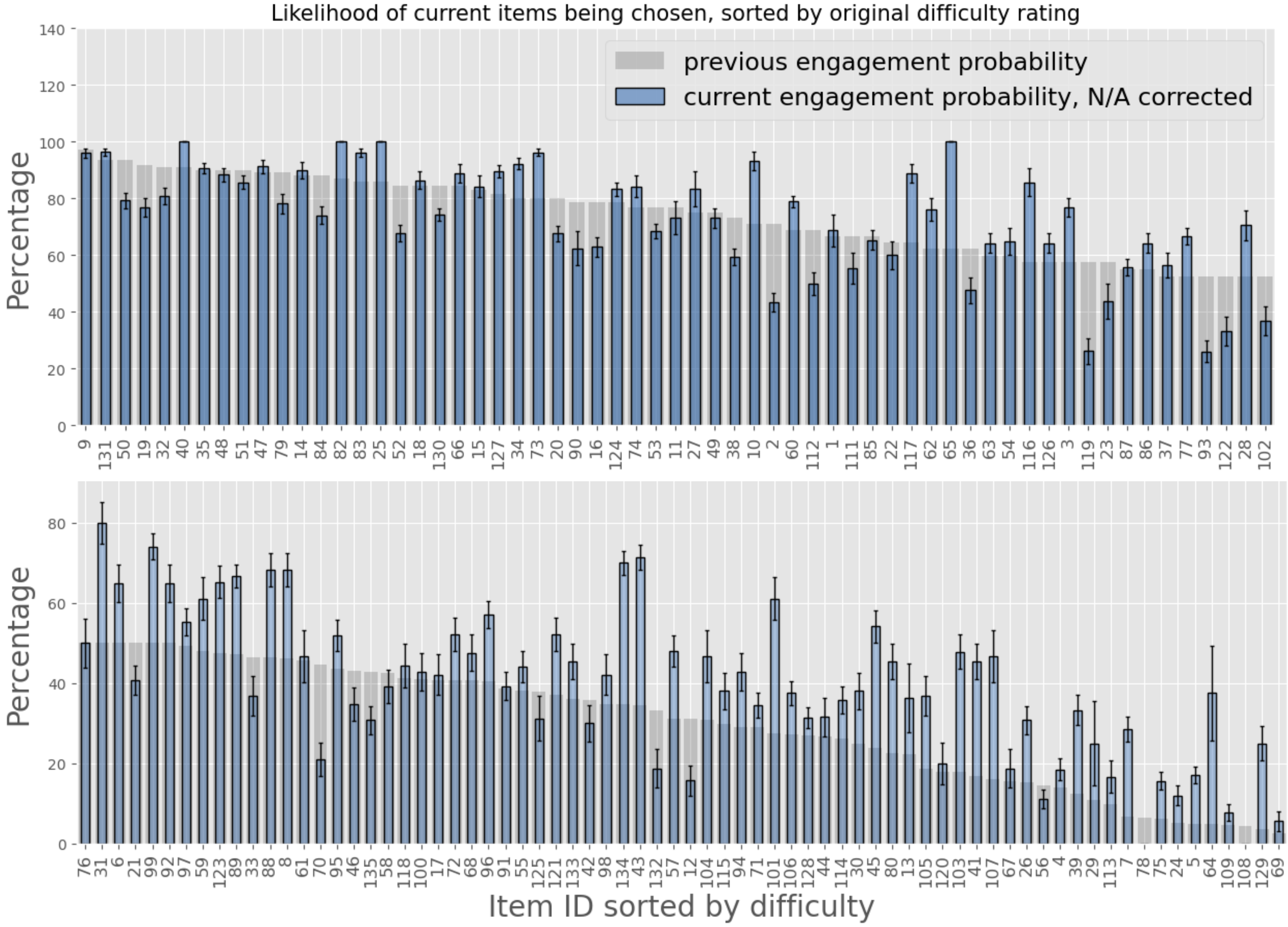
Item level results. We compared the original difficulty ratings of the measures, to the percentage of people who indicated that they already performed these measures in the 'current actions' screen. We did this merely visually, using the Python Panda's package. Because items were shown at random, with every person viewing 19 items, this means that all individual items were shown to 30 different people on average, with a minimum of 16 and a maximum of 45. In Figure 11, you can see the current engagement probability in blue, and the previous engagement probability in grey. This previous engagement probability is based on the probability that a person with an average ability of 0, would engage in a measure with a difficulty rating of δ_i, as calculated by the formula (3):

$$\textbf{Previous engagement probability} = \frac{e^{-\delta_i}}{1+e^{-\delta_i}} * 100\% \quad (3)$$

The current engagement probability is simply the number of times an item was shown, divided by the number of times someone indicated that they already performed that particular item (the latter minus the number of times someone indicated that an item was not applicable). For a full overview of all items, how often they were shown, and how often they were chosen in the current action screen and the new action screen, refer to Appendix F.

Figure 11

Previous and current engagement probabilities of measures, (Split into two rows to fit the page)



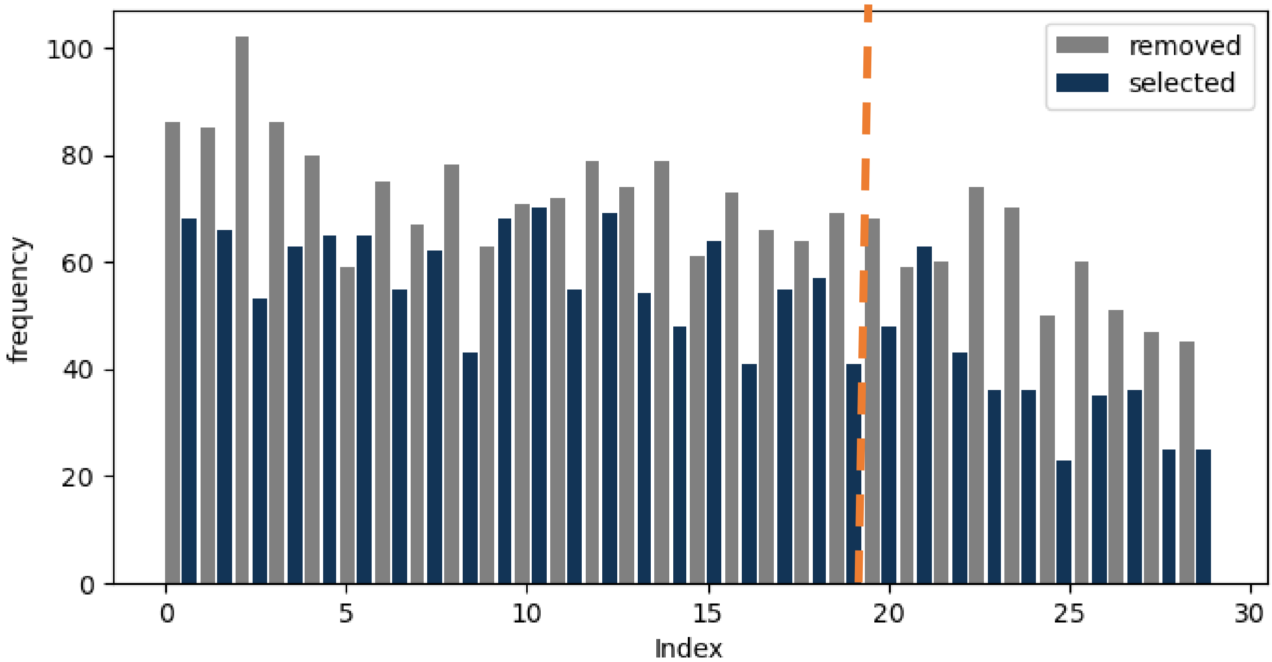
Bars represent 95% confidence intervals. Item 110 ("install a mini windmill") was not shown amongst current actions due to a programming error. The full figure can be found in Appendix G.

What stands out is that quite a few items, for example, item *70 (" Install a water-saving shower head")* and item 12 ("*Insulate your water heater so it feels cold to the touch."),* are chosen at a lower rate than previously (in 2018). On the other hand, some items, including 134 ("*Check the pressure in your boiler*."), 43 ("*Shower short(er)*"), and 45 ("*Turn off the tap while you soap yourself.")* are chosen at a much higher rate than we would expect based on the original difficulty ratings. However, we do not know the original confidence intervals for these difficulty ratings, though we expect them to be lower than the ones in our study, as items were previously shown at random to 304 participants at 25 items per person (Bams, 2018).

List indices. Below you see a graph of the position of items in the new item list, and how often they were selected or removed (Figure 12). Note that participants initially saw 20 measures (until the orange line), and only saw measures 21-29 if they removed other measures.

Figure 12

Item selection and removal by position in the recommendation list (Top to bottom of the list)



The orange dashed line depicts the 20th item cut-off; items right of this line were only shown when participants removed enough items.

Looking at the first 20 items (index 0 – 19), it seems that items are selected quite evenly across the list, but items at the beginning of the list seem to be more often removed.

Interactions. In the next section, we will discuss the data on participant level rather than on item level. We tested several interaction effects between demographics and system interactions (across all signposting and goal-setting conditions), which you can see in Table 5.

Table 5

Pairwise correlations for person characteristics. (N=202)

Variables	(1) NEP	(2) IMS	(3) Male	(4) Age	(5) kWh	(6) home	(7) En. lab	(8) Cur.	(9) new items	(10) goal
(1) NEP score	1.00									
(2) IMS core	**-0.25***	1.00								
(3) male	**-0.18***	0.04	1.00							
(4) age	0.02	**-0.22***	0.03	1.00						
(5) kWh Savings	-0.11	-0.03	0.00	0.04	1.00					
(6) homeowner	-0.13	0.01	0.13	**0.25***	0.12	1.00				
(7) Energy label	**-0.16***	-0.08	**0.15***	**0.21***	0.04	**0.21***	1.00			
(8) #Current items	-0.03	-0.11	0.12	**0.23***	**0.22***	0.04	**0.29***	1.00		
(9) #New Items	-0.11	0.10	-0.05	-0.07	**0.49***	-0.13	-0.01	**0.23***	1.00	
(10) goal Amount	-0.02	-0.07	0.03	-0.06	-0.11**	-0.05	0.13	0.10	**0.15***	1.00
(11) Education	0.07	-0.02	-0.05	**-0.22***	-0.04	-0.11	-0.10	-0.11	-0.03	0.01

** shows significance at p<.05 **Was significant with coefficient of 0.30 when considering only the goal condition*

From Table 5 we find that males score lower on the NEP scale. We also find that the NEP and IMS (Importance of money) scores are negatively correlated. Unlike in Starke (2019), we find no correlation between NEP scores and age. We do find a correlation between NEP score and gender, where males tend to score slightly lower on the NEP scale in our sample (again, unlike Starke 2019, p. 113). We find no correlation between being a homeowner and the number of current items chosen (ability level) or total chosen savings. We do find a correlation between age and ability level, with increasing age being correlated to performing more current items. Additionally, we find that people who perform more current items, also tend to pick more new items and select a higher total amount of savings. The latter is a general trend: people who choose more new items achieve a higher amount of savings from all selected items. We find no correlations between the height of the goal chosen and any of the other variables, except for the number of new items chosen. This means that people who score higher on the NEP scale or IMS scale, do not choose a higher goal.

4.1.2. Goal Conditions.

Our first hypothesis states that we expected guided goals to lead to higher energy savings than no goals. Guided goals meant that participants were able to choose between three goals: 600, 1200, and 1800 kWh. For the 97 participants in the goal condition, 10 chose the highest goal of 1800 kWh, 39 chose the medium goal of 1200 kWh, and 48 chose the low goal of 600 kWh. These goals might have been presented with a different signpost than kWh (Euro's or kg CO2), however, we will refer to them by the kWh values in this section. The mean savings and *SD* per goal condition can be seen in Table 6, in which the goal condition is further subdivided into the different goal amounts that people chose. We did a randomization check for ages and NEP scores between goal conditions and found no significant differences.

Table 6

Savings per goal condition, subdivided over self-selected goal amounts.

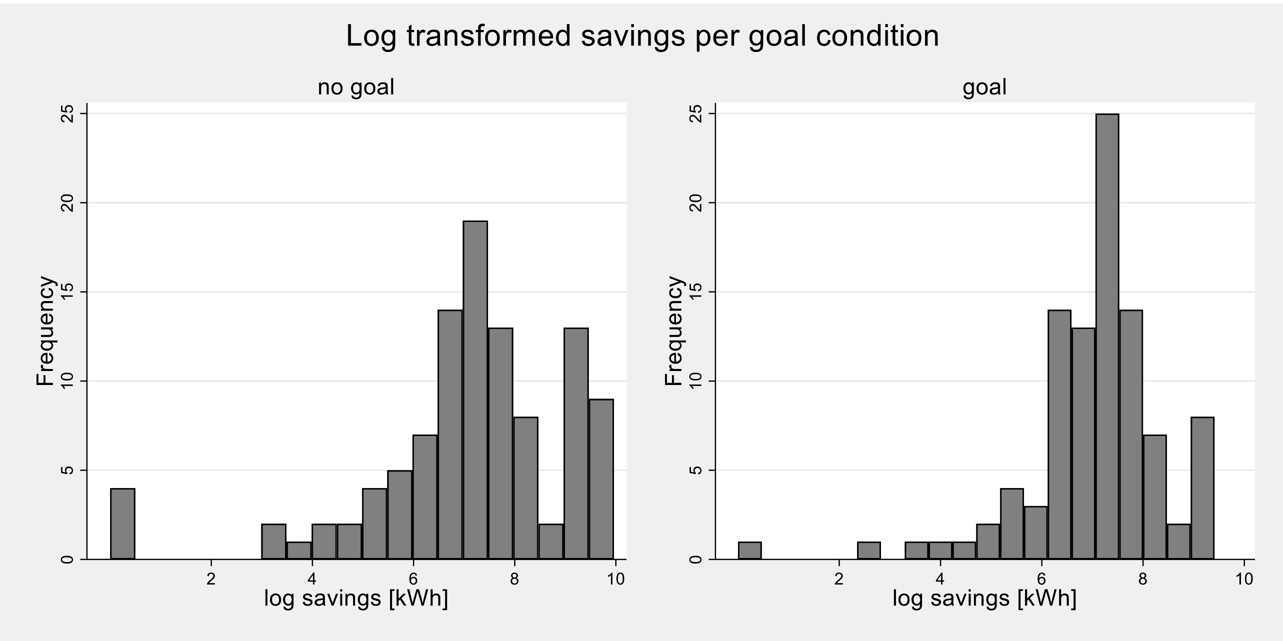
Goal Amount	Mean Savings (kWh)	*SD*	Median savings (kWh)	N	% goal reached
No goal	3880.4	5452.2	1470	105	/
Goal	2166.3	2624.6	1340	97	75.2%
Goal of 600	1503.8	2324.7	807.5	48	72.9%
Goal of 1200	2491.1	2546.8	1570	39	77.0%
Goal of 1800	4079	3317.3	3176	10	80%
Total	3057.27	4405.47	1362.5	202	/

From Table 4 it seems that higher selected goals seem to indeed lead to higher average savings (although the standard deviations are relatively high). We furthermore observe that in the goal-setting condition, 75% of participants reached their goal in terms of chosen savings, with a seemingly higher percentage reaching their goal for the high goal condition. Although this was only for ten observations.

Goal vs No goal. Because there was no normal distribution of savings for the no-goal and goal conditions, we performed a rank sum test to compare the savings in the two conditions. A rank sum test for kWh savings of selected measures, between the no-goal and goal condition, was not significant. Therefore, hypothesis 1 is not supported: there seems to be no difference in savings between the goal and no goal conditions. Though, this might change if we do find an effect in the follow-up study on actual savings, because that would indicate that goal setting eventually does lead to a saving differential for actual behavior, rather than what participants say they will do. For the effects on savings, we will therefore adhere to a corrected p-value of 0.025 since we perform two tests. You can see the distributions of savings for both goal-setting conditions in Figure 13, which indeed seem quite similar.

Figure 13

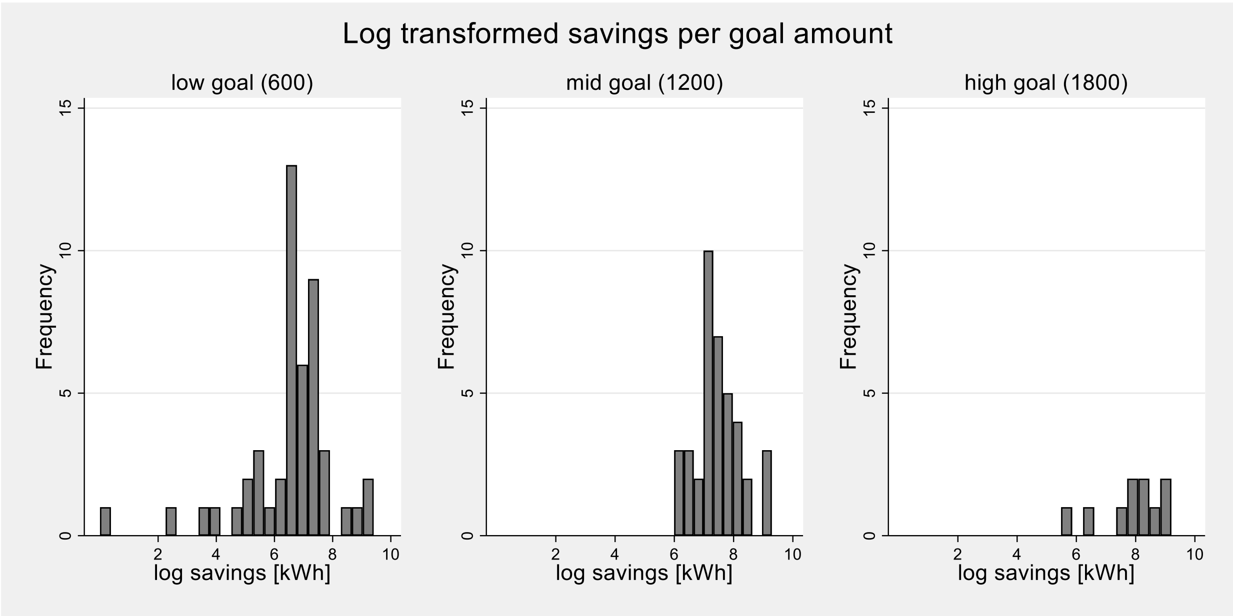
(Log-transformed) Savings per goal condition



Goal Amounts. We furthermore compared the savings between the three goal amounts, as graphed below in Figure 14. This was not part of any hypothesis but might still give some insight into what extent people are motivated to reach their self-chosen goal. What stands out is that in the 600-kWh condition, there are more participants with very low chosen savings. This might indicate that those without much motivation to save energy, chose this goal.

Figure 14

(Log transformed) Savings per goal amount



To examine this difference, we performed a Kruskal-Wallis equality-of-populations rank test between the chosen goals, because the logarithmic transformation did not lead to normal distributions for all conditions. This test was significant with $\chi^2 = 19.43$ at $p<0.001$, indicating that the differences in chosen savings between goal amounts were significant (of which the averages could be seen in Table 6).

Number of items per goal. In addition to the total amount of savings, we looked at the number of items chosen per condition and self-chosen goal. Again, this was additional and not necessarily related to the research questions or hypotheses. In Table 7 you can see the number of items chosen and the number of items removed per condition. From this table, it seems that people who chose the goal condition of 1200 kWh, chose the highest number of items, followed by the 1800 kWh condition and finally the 600 kWh condition.

Table 7

System interactions for goal conditions

Goal Amount	mean # of items chosen	*SD*	mean # of removed items	*SD*	Freq.
No goal	7.4	4.2	11.1	6.3	105
Goal	7.8	3.3	9.4	6.7	97
Goal of 600	6.3	2.8	8.5	6.2	48
Goal of 1200	9.7	2.9	10.2	6.9	39
Goal of 1800	7.8	3.6	10.6	8.2	10
Total	7.58	4.29	10.26	6.51	202

The median number of chosen items was 7, in both the goal and no-goal condition. A t-test showed no significant difference in number of items chosen between the goal and no-goal condition (*t(200)=.71, p=*.48). An ANOVA for the number of new items between the chosen goal amounts of 600, 1200, and 1800 was significant ($F(2,94)=14.38$, $p<0.001$), indicating that there was indeed a difference between the number of items chosen per self-chosen goal amount as seen in table 7, with people choosing the most items in the 1200kWh condition with an average of 9.7 items.

To understand the interaction between on the one hand seeing that higher chosen goals lead to higher savings, and on the other hand seeing that those who chose a higher goal amount, also chose fewer items, we compared the average kWh per item across chosen goal amounts. We first performed a logarithmic transformation of the kWh/item variable to ensure a normal distribution in all three conditions. The analysis of variance (ANOVA) for log savings between the goal amounts was significant ($F(2,93)=5.58$, $p<0.01$), with the average kWh savings per item in the 600kWh condition being equal to 234 (*SD* of 340), the average saving being 257 (*SD* = 231) for the 1200kWh condition, and 541 (*SD* = 516) for the 1800kWh condition. This would mean that in the 1800 condition, people choose items with higher savings (albeit based on just ten observations).

Because goals were signposted, we compared the chosen goal amounts between signpost conditions, to see if people in different signpost conditions might have chosen different goal levels. A Fisher's exact test between all three signpost conditions and all three goal conditions was not significant χ^2 (four degrees of freedom) = 4.19, with p=.38, indicating that signposts did not influence which goals people chose.

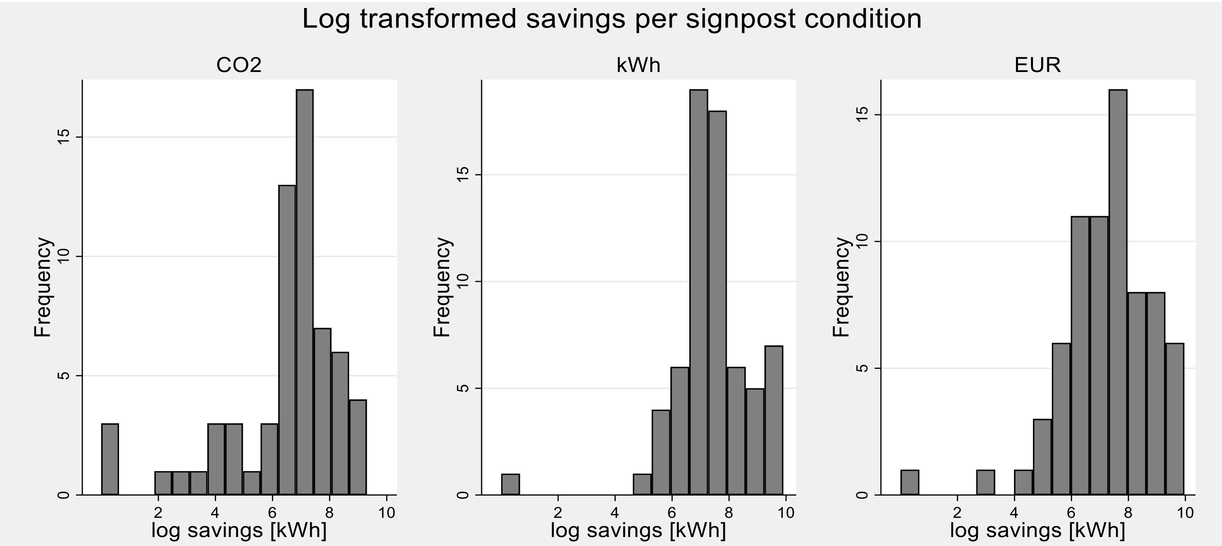
Six one-way ANOVA's were furthermore used to compare NEP scores and IMS scores (2*) between chosen goal amounts, for all three signpost conditions (3*). This was done to see if user values influenced chosen goal amounts in particular signpost conditions. This resulted in one significant interaction effect: In the Euro condition, NEP scores influenced the goal amount chosen ($F(3, 68)$ = 4.05, $p<0.05$). We found that the average NEP score in the Euro condition (on a scale of -3 to 3), for the 11 participants who chose the 600 kWh goal, was .38 (*SD* .57), for the 15 participants who choose the 1200 kWh goal, this was higher with an average of 1.03 (*SD* .48) and for the 5 participants who chose the 1800 kWh goal, the NEP score was again slightly lower at .92 (*SD* = .38). For the other signpost conditions, and for the IMS scores, we do not observe such interaction effects. However, as the goal amount is not considered to be an independent variable, this interaction should not have a large impact on our results.

4.1.3. Signpost Conditions

Next, we compared the chosen savings per signpost condition. We first checked for sufficient randomization of ages and NEP scores for all three signpost conditions and found no significant differences. In Figure 15 you can see the log-transformed savings per signpost condition. What stands out is that the CO2 signpost seemingly results in a larger portion of lower savings, and that the EUR signpost seems to result in a more normal distribution of savings. The peaks at ~7 for the CO2 and kWh signpost conditions, correspond to savings around $e^7 \approx$ 1100 kWh (Stata defaults to natural log), which might correspond to the goal amount of 1200 kWh.

Figure 15

(log-transformed) Savings per signpost condition.



The hypotheses on signposts will be discussed later, in the SEM model section, as we only hypothesized effects of signposts in interaction with personal values, and not signposts by themselves. However, as you can see in Table 8, it seems that the savings of selected measures were highest in the Euro and kWh conditions, and lower in the CO2 condition.

Table 8

Savings for each signpost condition

Signpost	Mean Savings (kWh)	*SD*	Median	Freq.
CO2	1781.5	2553.6	980	63
EUR	3619.9	4842.7	1592.5	72
KWh	3652.1	5049.2	1553	67
Total	3057.3	4405.5	1350	202

We tested if this difference was indeed significant. Because no transformations worked to render the data normally distributed, we performed a Kruskal-Wallis equality-of-populations rank test between the signposts. This test was significant with $\chi^2 = 10.11$ and $p<0.001$, indicating that there was indeed a difference in savings between signpost conditions with lower savings for the CO2 condition.

We also looked at the total number of items chosen between signpost conditions. As you can see in Table 9, fewer items seemed to be chosen, and more seemed to be removed, in the CO2 condition.

Table 9

System interactions per signpost condition

signpost	Mean # of new items	*SD*	Mean # of removed items	*SD*	Freq.
CO2	6.2	3.4	11.0	7.0	63
EUR	8.1	3.8	9.5	6.2	72
kWh	8.3	3.8	10.3	6.3	67
Total	7.6	3.8	10.3	6.5	202

A one-way ANOVA for the number of items between signpost conditions was indeed significant ($F(2, 199)$, 6.09, $p<0.01$). We did not formally compare removal rates.

4.1.4. Direct Effects from Signposts and Goals on Self-Efficacy

To understand how system manipulations and interactions with user values influence energy-saving self-efficacy in users, we performed a robust linear regression (Table 10). For this, we used the self-efficacy score as computed by Stata. (Note: whenever '*self-efficacy*' is mentioned in the below section and in the SEM model, '*energy-saving self-efficacy*' is implied).

Table 10

Robust linear regression predicting the self-efficacy score based on experimental conditions and NEP and IMS scores.

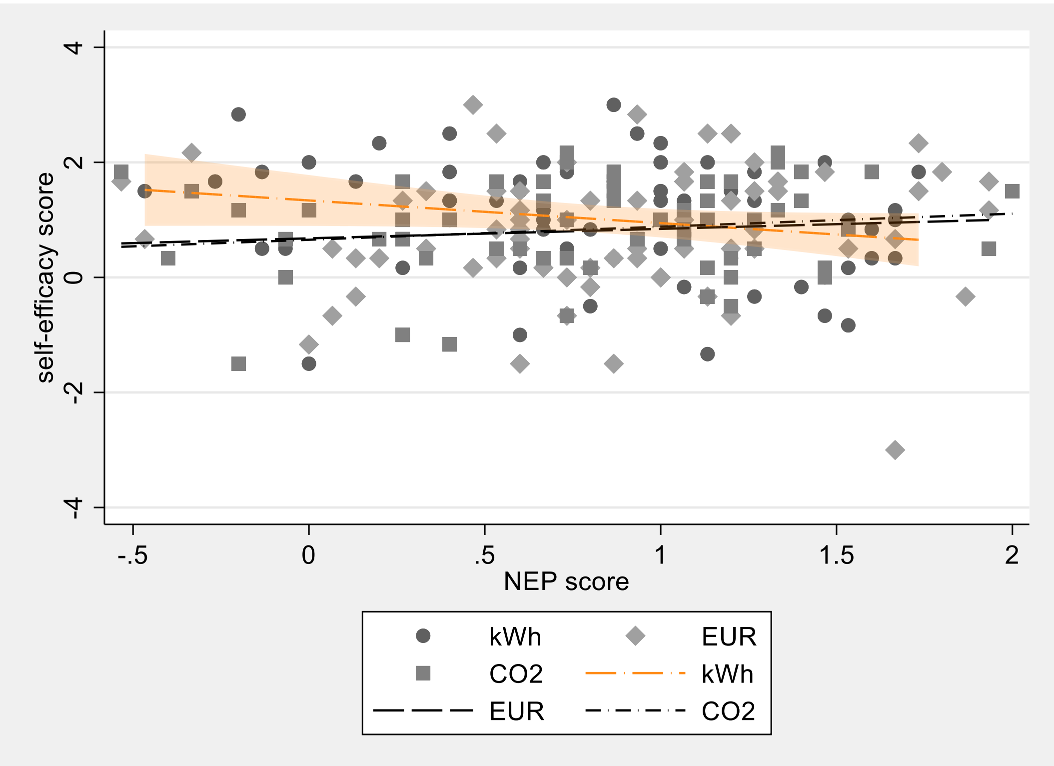
Variables	Self-efficacy Coef.	β	*SE*	95% *CI*
IMS Score	-0.01	-0.04	(0.16)	[-0.37; 0.26]
Signpost				
CO2	-0.94*	-0.45*	(0.39)	[-1.70; -0.18]
EUR	-0.83	-0.41	(0.44)	[-1.70; 0.04]
Signpost # IMS score				
CO2	0.010	0.04	(0.22)	[-0.34; 0.52]
EUR	-0.01	-0.00	(0.29)	[-0.59; 0.56]
NEP score	-0.47	-0.26	(0.24)	[-0.94; 0.01]
Signpost # NEP score				
CO2	0.70*	0.35*	(0.32)	[0.07; 1.33]
EUR	0.61	0.33	(0.36)	[-0.09; 1.32]
Goal condition				
No Goal	-0.26	-0.13	(0.26)	[-0.76; 0.25]
Goal condition#signpost				
No Goal # CO2	0.32	0.12	(0.34)	[-0.34; 0.99]
No Goal # EUR	0.24	0.098	(0.37)	[-0.49; 0.96]
Constant	1.55**		(0.32)	[0.92; 2.17]
Observations	202			
R-squared	0.039			

*** p<0.01, * p<0.05*

This regression shows us that there is a negative effect of the CO2 signpost on energy-saving self-efficacy as compared to the kWh signpost (β=-.45, *p*<.05), while we see a positive interaction effect of the NEP score on self-efficacy in the CO2 condition (β=.35, $p<.05$), as compared to the kWh signpost. This last finding is in line with our 13th hypothesis, which stated that CO2 signposting, as compared to kWh signposting, would result in stronger environmental self-efficacy for increasing strength of pro-environmental values. However, we expected that this effect would be caused by 'signposting' as previously observed by Ungemach et al.; We expected that kWh would serve as a 'neutral' baseline, and the CO2 signpost would show a clear difference for lower and higher NEP scores, as pro-environmental values would be activated by this CO2 signpost. However, when looking at the graph in Figure 16, we see that rather than an upward slope for the CO2 condition, we observe only a downward slope for the kWh condition, in which people who have higher nep scores, report lower feelings of energy-saving self-efficacy. So, while the hypothesized effect is observed in the right direction, the underlying causes do not seem to be in line with the rationale we proposed in the introduction.

Figure 16

The effect of NEP score on self-efficacy for the three signpost conditions



The orange shaded area depicts the 95% confidence interval of the kWh signpost.

4.1.5. Direct Effects from Signpost and Goals on Choice Satisfaction

Additionally, we looked at the effect of user values and system manipulations on choice satisfaction with a robust linear regression, as can be seen in Table 11. From this regression, we find that the CO2 signpost leads to decreased choice satisfaction as compared to the kWh signpost (β = -0.59, $p<0.01$). We also find that there is a positive correlation between NEP scores and choice satisfaction in the CO2 condition, as compared to in the kWh condition (β = .30, $p<0.01$). This lends support to our 10^{th} hypothesis; however, we face the same issue as with the self-efficacy scale as explained below.

Table 11

Robust linear regression predicting choice satisfaction score based on experimental conditions and NEP and IMS scores.

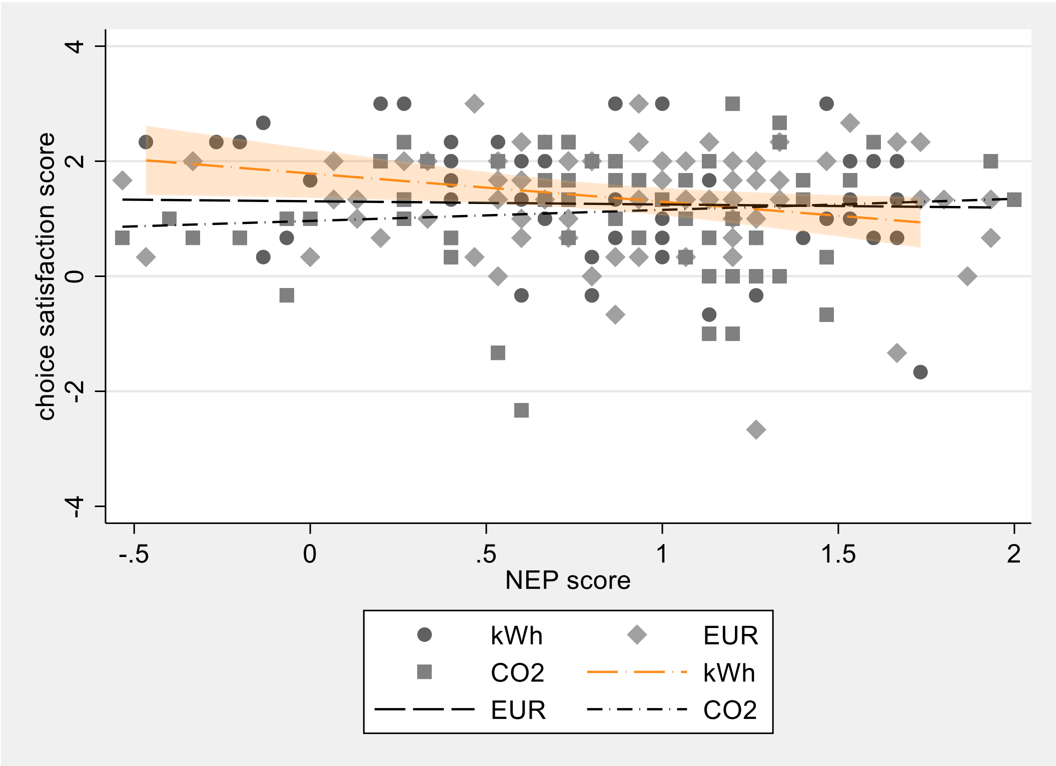
Variables	Choice satisfaction Coef.	β	*SE*	95% *CI*
IMS Score	-0.09	-0.07	(0.16)	[-0.41; 0.22]
Signpost				
CO2	-1.25**	-0.59**	(0.37)	[-1.98; -0.52]
EUR	-0.59	-0.29	(0.40)	[-1.37; 0.19]
Signpost # IMS score				
CO2	0.34	0.16	(0.21)	[-0.08; 0.76]
EUR	0.23	0.10	(0.23)	[-0.23; 0.69]
NEP score	-0.55*	-0.31*	(0.24)	[-1.02; -0.07]
Signpost # NEP score				
CO2	0.78**	0.30**	(0.29)	[0.21; 1.35]
EUR	0.56	0.30	(0.32)	[-0.07; 1.18]
Goal condition				
No Goal	-0.14	-0.07	(0.26)	[-0.65; 0.37]
Goal condition # signpost				
No Goal # CO2	0.51	0.18	(0.36)	[-0.20; 1.22]
No goal # EUR	-0.07	-0.028	(0.33)	[-0.72; 0.59]
Constant	1.93**		(0.31)	[1.32; 2.54]
Observations	202			
R-squared	0.069			

*** p<0.01, * p<0.05*

When looking at the graph of nep scores and choice satisfaction for the different signpost conditions (Figure 17), we observe a negative correlation for the kWh signpost with increasing nep score, while there does not seem to be a strong effect for the other two signposts with increasing NEP score. We originally expected an upward slope for the CO2 condition and a horizontal line for the kWh signpost. We find no effects of the importance of money scores and the monetary signpost on choice satisfaction.

Figure 17

The effect of NEP score on choice satisfaction for the three signpost conditions



The orange shaded area depicts the 95% confidence interval of the kWh signpost.

4.1.6. Direct Effects from Signposts and Goals on Savings

A linear regression with all signposts, goal conditions, and interactions between values and signposts (Table 11), shows that a CO2 signpost leads to higher savings than a kWh signpost (β = -0.38, $p<0.05$). We do not find an effect of goal setting on savings, nor an interaction effect of goal conditions and signposts on savings. There is no interaction effect between NEP scores and signposts as we saw with self-efficacy and choice satisfaction scales.

Table 11

Robust linear regression predicting the log savings based on experimental conditions and NEP and IMS scores.

Variables	Log savings (kWh) Coef.	β	*SE*	95% *CI*
IMS score	-0.25	-0.10	(0.23)	[-0.70; 0.20]
Signpost				
CO2	-1.44*	-0.38*	(0.69)	[-2.80; -0.08]
EUR	0.07	0.019	(0.670)	[-1.30; 1.44]
Signpost # IMS score				
CO2	0.72	0.18	(0.39)	[-0.05; 1.49]
EUR	0.45	0.11	(0.36)	[-0.25; 1.15]
NEP score	-0.80	-0.25	(0.51)	[-1.81; 0.21]
Signpost# NEP score				
CO2	0.38	0.10	(0.60)	[-0.81; 1.56]
EUR	0.16	0.05	(0.65)	[-1.13; 1.45]
Goal condition				
No Goal	0.26	0.07	(0.42)	[-0.57; 1.08]
Goal condition # signpost				
No Goal # CO2	-0.14	-0.03	(0.67)	[-1.45; 1.18]
No Goal # EUR	-0.75	-0.17	(0.58)	[-1.89; 0.38]
Constant	8.01**		(0.55)	[6.91; 9.10]
Observations	202			
R-squared	*0.132**			

*** p<0.01, * p<0.05*

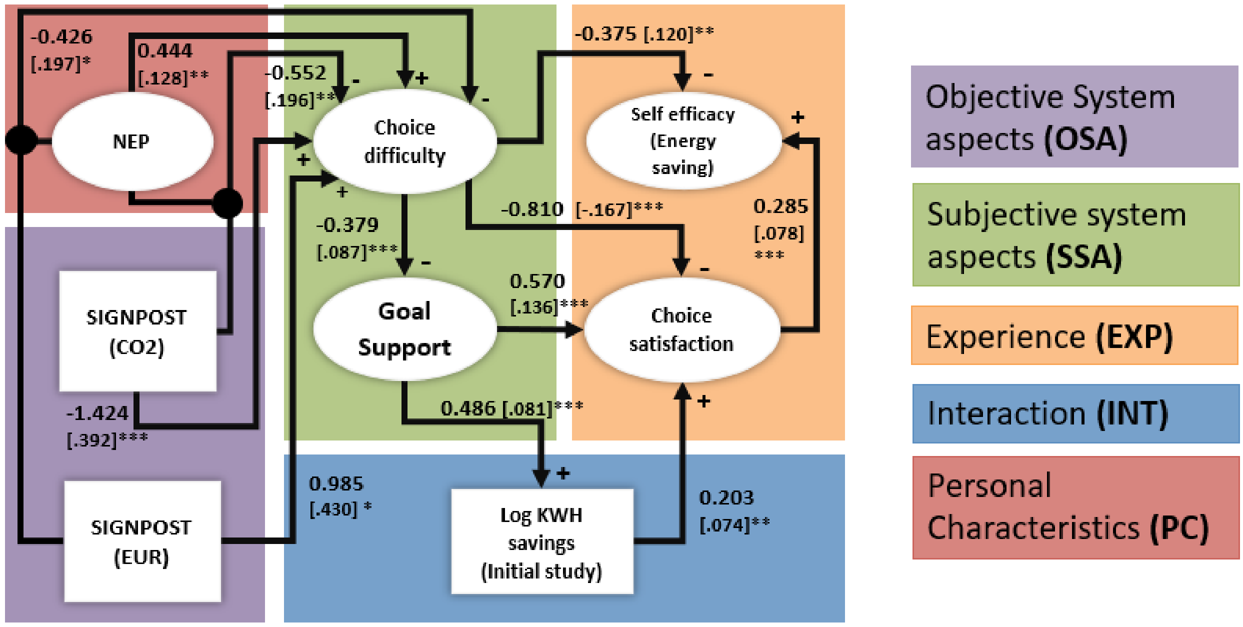
4.1.7. Factor Analysis and Structural Equation Model (SEM)

To understand the relationship between these system manipulations on the one hand, the user interactions, and experiences on the other, and how they are mediated by how users perceive the system, we performed an exploratory and confirmatory factor analysis on the various user experience questions (Refer to measures section 3.4.1.). After this, the observed variables (system manipulations) and latent factors resulting from the factor analysis, were analyzed in a structural equation model.

Structural equation model. We devised a structural equation model with mPlus, to understand the relationship between objective system aspects (Experimental conditions), subjective system aspects (Choice difficulty and goal support), experience factors (Energy saving self-efficacy and choice satisfaction), and interactions (Savings). We chose to focus on savings, rather than for example the number of items, because eventually, the total savings will have a larger impact on both the environment and energy costs than the number of measures performed. We started with a saturated model with all hypothesized connections and from there removed insignificant paths. The results and standard errors can be seen in Figure 18, and the mPlus outputs can be found in Appendix H (Please note that the choice difficulty scale is reversed in this output). The baseline in this model is the kWh signpost.

Figure 18

Structural equation model (SEM) for the initial study. Coefficients are depicted as numbers on the arrows, and standard errors in brackets. (Design in line with evaluation framework of Knijnenburg and Willemsen (2015))



**** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$*

Based on the guidelines in the handbook *'Evaluating Recommender Systems with User Experiments'* by Knijnenburg and Willemsen (2015), the fit of our SEM model was good, with χ^2 (N=202) of 255.00 (p<.01), RMSE of .040 (<.05) (90% *CI* between .025 and .052), comparative fit index CFI of .981 (>.96) and Tucker-Lewis Index TLI of .979 (>.95).

We find several correlations between signposts and latent variables, some of which are mediated by pro-environmental values (NEP score). We find no direct effects from signposts and interactions on outcome variables (Savings, choice satisfaction, and self-efficacy). From the model in Figure 18, we observe that all possible effects from signposts and values on these outcome variables are mediated by choice difficulty. Effects on savings are furthermore mediated by goal support.

The importance of money score was not found to influence savings or other factors, thus not finding support for hypotheses 14, 15, 16, 17, and 18, which predicted effects of monetary values on various latent variables, through interactions with financial values. Note that the interaction with IMS and the

CO_2 signpost were omitted here because it was not a hypothesized effect, and we performed the SEM before we performed the regressions. Additionally, discriminant validity of choice satisfaction (AVE .66) is not maintained, because choice satisfaction is better explained by choice difficulty (β = -.81, p<.001) than by its own questions. However, we decided to keep this factor in the model because choice difficulty is generally accepted to be a subjective system aspect, whereas choice satisfaction is usually referred to as an experience outcome as per the user-centric evaluation framework by Knijnenburg and Willemsen (2015). Therefore, merging these into one construct would not be entirely logical, and it would make it harder to gain insight into the relationship between these categories of latent variables.

Below, we will first discuss how system characteristics affect subjective system aspects (Choice difficulty and goal support), which are found in the green area (middle) of Figure 18, and how these aspects influence each other. Then, we will discuss the outcome variables of the model (interaction outcomes, which concern the total chosen savings, and experience outcomes, referring to self-efficacy and the choice satisfaction factors). These can be found in the (bottom) left part of Figure 18, in the blue and orange area. Along with these outcome variables, we will discuss the indirect effects from signposts via the subjective system aspects on these variables, and how these connect to the hypotheses we had before.

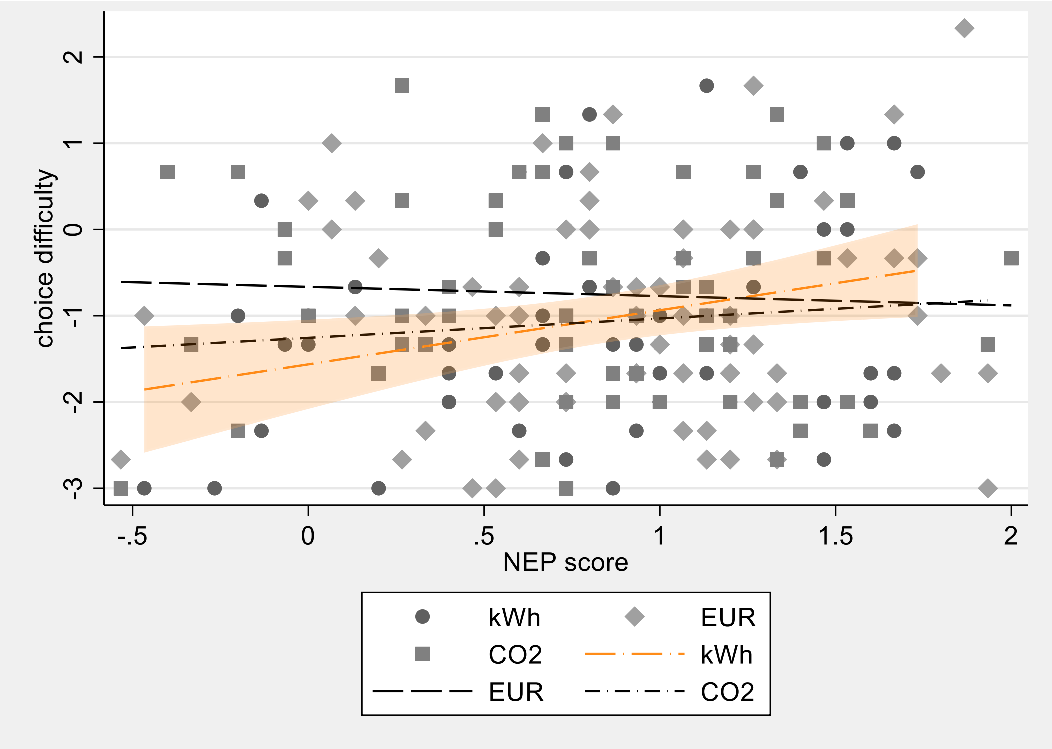
Choice Difficulty. We find a positive effect of NEP score on choice difficulty (β = .44, p<.01), a positive effect from the Euro signpost on choice difficulty (β = .99, p<.05), and a negative effect from the CO_2 signpost on choice difficulty (β =-1.42, p<.001). We furthermore find a negative interaction effect of the CO_2 signpost and NEP score on choice difficulty (β = -.55, p<.01), and a negative interaction effect of the Euro signpost and NEP score on choice difficulty (β =-.426, p<.05), indicating that those with higher NEP scores report less choice difficulty in the CO_2 and Euro conditions as compared to the kWh condition. The former is in line with hypothesis 9, that CO_2 signposting would result in lower reduced choice difficulty as compared to the kWh signpost, for increasing NEP scores.

To better understand the effect of NEP scores on choice difficulty in the three signpost conditions, we plotted these correlations as can be seen in Figure 19 below. The orange trend-line depicts the linear relation between NEP score and choice difficulty for those who were presented with the kWh signpost (with a

95% confidence interval). Here, we see that in the kWh condition, those with a higher NEP score experience greater choice difficulty than those with a lower NEP score. However, we do not see a clear trend for the other two signpost conditions; we do not see the expected downward trend for the CO2 condition. Therefore, as we saw previously with this interaction effect on self-efficacy and choice satisfaction (regressions), we do find support for the hypothesis, although the effects do not manifest in the way we expected them to.

Figure 19

Effect of NEP score on choice difficulty in different signpost conditions.



The orange shaded area depicts the 95% confidence interval of the kWh signpost.

Goal Support. We furthermore see various interactions between latent constructs. Choice difficulty is negatively related to goal support (β = -.38, p<.001) indicating that lower choice difficulty results in higher goal support. This provides partial support for our 29th hypothesis, that a decrease in choice difficulty would lead to increased system satisfaction, a construct that was merged with the goal support scale due to high covariances. Increased goal support then leads to higher savings (β = .49, p<0.001), which is in line with our 28th hypothesis.

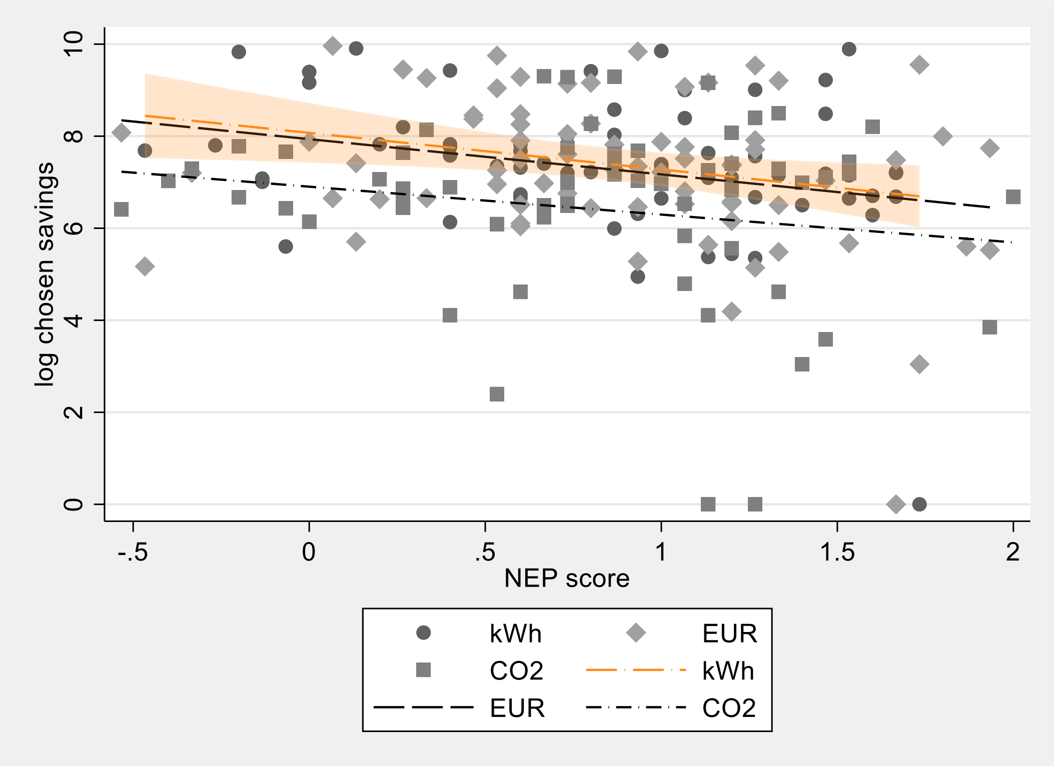
Savings. The indirect effect of the CO2 signpost on savings is β = -.26 (p<.01), via choice difficulty and goal support. Thus, the reason people save less in the CO2 condition, as compared to the kWh condition is that they have more difficulty choosing between items, and this in turn leads them to feel less supported in their energy-saving pursuits, which eventually leads to lower savings.

The same can be seen for the Euro signpost, albeit this effect is slightly weaker (β = -.18, p=.05, via the same pathway). Furthermore, there is a direct, though close to zero, negative indirect effect from NEP score on savings via this same pathway (β = -.082, p<.01), indicating that those with a higher NEP score choose slightly less savings because they experience greater choice difficulty and in turn lower goal support.

We furthermore see a small indirect interaction effect on savings, in which the interaction of NEP and CO2 leads to higher savings (β = .10, p<.01), meaning that increasing NEP score leads to slightly higher savings in the CO2 condition as compared to the kWh condition, and this is caused by a decrease in choice difficulty and increase in goal support. This is in line with hypothesis 8, which stated that the CO2 signpost would lead to higher savings as compared to the kWh signpost, for increasing strength of pro-environmental values. This is something we did not observe in the regression in section 4.1.6, possibly because more variables (IMS scores and goal conditions) were included there that were not significant. However, even if we correct for this double testing, this interaction finding would still remain significant (With p being smaller than .025). In the figure below, you can see this interaction effect.

Figure 20

Effect of NEP score on chosen savings for different signpost conditions.



The orange shaded area depicts the 95% confidence interval of the kWh signpost.

We mainly see in Figure 20, that all lines have a downward slope, with the kWh line being higher than all the other lines across Nep scores. The CO2 has a slightly less steep download slope, resulting in this observed effect.

Self-efficacy. We see a negative correlation between choice difficulty and self-efficacy (β = -.38, *p*<0.01) indicating that people who experience less choice difficulty, report higher energy-saving self-efficacy. This is in line with our 19th hypothesis. We do need to mention here that, while our system does have an effect on self-efficacy, existing feelings of energy-saving self-efficacy might also have an impact on choice difficulty and thus the directionality of this effect might be twofold: someone who is already more confident in their energy saving ability, might find it less difficult to choose between items.

Additionally, we observe a positive effect of choice satisfaction on self-efficacy (β = .29, *p*<.001), which is in line with our 31st hypothesis. We do not observe a direct effect of goal support on self-efficacy, thus not finding support for our 26th hypothesis that increased goal support would lead to increased energy-saving self-efficacy.

We furthermore observe various indirect effects on self-efficacy. We observe a total negative indirect effect of the CO2 signpost on self-efficacy $\beta = -.97$ ($p<.01$), via 4 different pathways: The CO2 signpost leads to lower self-efficacy via increased choice difficulty ($\beta = -.53$, $p<.05$), increased choice difficulty and decreased choice satisfaction ($\beta = -.33$, $p<.01$), and lastly two rather weak effects via increased choice difficulty, decreased goal support and decreased choice satisfaction ($\beta = -.09$, $p<.05$) and via the same pathway with log savings included between goal support and choice satisfaction ($\beta = -.02$, $p<0.05$). We see here that the longer the pathway, the weaker the indirect effect becomes, and we observe the strongest indirect effect via choice difficulty, indicating that the CO2 signpost leads to lower savings primarily because people experience greater choice difficulty in this condition as compared to the kWh condition, and secondarily because they in turn experience lower choice satisfaction and this leads to lower self-efficacy.

We also observe a negative indirect effect from the Euro signpost on self-efficacy ($\beta = -.67$, p<.05), of which the significant pathway goes via increased choice difficulty and decreased choice satisfaction ($\beta = -.23$, $p<.05$).

Additionally, we observe a negative indirect effect from the NEP score on self-efficacy ($\beta = -.301$, $p<.01$), via increased choice difficulty, ($\beta = -.17$, $p<.05$), increased choice difficulty and decreased choice satisfaction ($\beta = -.10$, $p<.01$). The latter pathway can be extended via decreased goal support in the same way as with the CO2 signpost ($\beta = -.02$, $p<.05$) and further via decreased savings ($\beta = -.01$, $p<.05$).

Then, we observe two indirect interaction effects of signposts and NEP scores on self-efficacy, of which the first is an indirect positive effect of the NEP score and CO2 signpost on self-efficacy ($\beta = -.37$, $p<.05$), mainly via decreased choice difficulty ($\beta = -.21$, $p<.05$), and via decreased choice difficulty and increased choice satisfaction ($\beta = .13$, $p<0.05$). This provides further support, and a further explanation, for our 13th hypothesis that the CO2 signpost would lead to stronger feelings of self-efficacy than the kWh signpost, for increasing NEP scores (The regression in section 4.1.4. also supported this hypothesis). This means that the increase in self-efficacy for increasing NEP scores, for the CO2 signpost, can be explained by the fact that those with a higher NEP score experience less choice difficulty when presented with a CO2 signpost as compared to the kWh signpost, which in turn leads to them feeling more strongly supported in reaching their energy-saving goals (goal support), and this increases their feelings of energy-saving self-efficacy.

The second indirect effect on self efficacy, is an indirect positive effect of the EUR signpost and NEP score on self-efficacy (β = -.29, p<.041), of which the significant pathway goes via decreased choice difficulty and increased choice satisfaction (β =.10, p<.05). This is again something we did not necessarily expect beforehand, as NEP scores were only thought to be activated by the CO2 signpost.

These two indirect interaction effects indicate that, although the CO2 and Euro signpost lead to worse self-efficacy in isolation as compared to the kWh signpost, an increasing NEP score mitigates these effects to an extent.

Choice Satisfaction. For the outcome measure of choice satisfaction, see a positive effect of savings (β = .20, p<.01), indicating that those who selected a higher total number of energy savings, feel more satisfied with these choices, which is in line with our 30th hypothesis. Note that for the goal condition, higher savings means a higher likelihood that people have in fact reached their goal, which might impact this relationship. We furthermore see that increased goal support leads to stronger feelings of choice satisfaction (β = .57, p<.001), which is in line with our 27th hypothesis.

Several indirect effects from signpost conditions and interactions on choice satisfaction can be observed; The CO2 signpost leads to lower choice satisfaction (β = -1.51, p<.01), via increased choice difficulty (β = 1.15, p<.01), and increased choice difficulty and decreased goal support (β = .31, p<.01), indicating that a CO2 signpost leads to decreased choice satisfaction because people experience higher choice difficulty as compared to a kWh signpost, and also because they in turn experience lower goal support, which both lead to lower feelings of choice satisfaction. We see the same indirect effect for the Euro signpost (β = -1.05, p<.05), and this indirect effect on self-efficacy goes mainly via increased choice difficulty (β = -.80, p<.05). We also see this negative effect for the NEP score on choice satisfaction; (-.47, p<0.01), which mainly goes via increased choice difficulty (β = -.36, p<.01) and increased choice difficulty and decreased goal support (β = -.10, p<.01).

We observe several interaction effects on choice satisfaction; an increasing NEP score leads to increased choice satisfaction in the CO2 condition (β = .59, p<.05), which is fully explained by indirect effects, of which the significant pathways are via choice difficulty (β = .45, p<.05) and choice difficulty and goal support (β = .12, p<.05). Indicating that an increasing NEP score leads to decreasing choice satisfaction in the CO2 condition as compared to the kWh condition, because of lower choice difficulty and stronger feelings of goal-support. This overall effect is in line with the 10th hypothesis, also seen in the regression in section 4.1.5. We observe the same interaction with NEP for the Euro signpost, (β = .45, p<.05), which is primarily explained by the pathway via choice difficulty (β = .35, p=.05). We did not expect these and other interactions with Euro signposts and NEP scores, as explained before in the regression section, but they could thus be explained in the same way as the interaction between CO2 signposts and NEP scores; apparently, increasing NEP score results in higher choice satisfaction in the Euro condition, because people with higher NEP scores experience less choice difficulty in this condition, as compared to in the kWh condition. We showed before that these effects are mostly due to a negative effect of NEP on choice satisfaction in the kWh condition, rather than a positive effect of NEP score on choice satisfaction in the Euro or CO2 conditions (Figure 17).

4.2. Follow-Up Results

In the follow-up study, participants were asked, approximately four weeks after the original study, which measures they actually performed out of their chosen measures. 'Actually performed' refers to the items participants indicated as 'still performing', and not the items participants started with or that they planned to do. Including these other categories made no difference in the significance of the results of this section. In the follow-up interface, participants were again presented with the same signposts as before. In total**,** 170 Participants completed the follow-up study by June 23rd, of which 84 were from Archie, one was external, and 85 were from Prolific. In total, this amounted to 80% of the original sample**.** Of these, 56% were female, 42% were male, 2% other, and they had a mean age of 30.0 (*SD* 13.39). Again, 74% of participants were non-homeowners, the same as in the initial study. Energy labels and dwelling types were also similar, refer to Appendix I for the full overview. Building years and education levels were also almost identical to the original sample. We did not observe any direct demographic-based bias in our follow-up sample compared to the initial sample, except that the mean chosen savings were 3057 kWh in the initial study, whereas the initial chosen savings of the follow-up sample averaged at 3382 kWh; thus, this reduced sample had slightly higher savings in the initial study than the overall sample.

We again removed several participants for various reasons. Four participants who did not choose any items in the initial study were removed, as were two participants who did not indicate for all the items presented whether they still performed them or not; something that was instructed. We furthermore excluded the outliers from the initial study who had shown little variation in scale answers (two participants), completed the study in 6 minutes while clicking all the items presented (one participant), and one participant who did not live in the Netherlands. We did however include the participants who were previously removed for choosing an unrealistic number of items (above 16, 2 *SD* above the mean), as we saw that these participants indeed indicated to be performing a lot of items after four weeks, yet at a more reasonable level of max. ten items. We did not remove participants who still performed more than the average amount of items + 2*SD*, as this would mean removing everyone who performed more than six items, which we deemed a reasonable amount. After removing outliers, 160 participants remained. The gender ratios of this reduced sample were the same, and the mean age was nearly the same as well at 29.7 (*SD* 13.1).

4.2.1. Descriptive Statistics.

We found that participants, on average, still performed 2.1 measures (*SD* 1.9), with total average savings of 315.8 kWh (*SD* 538.2 kWh). The maximum count of measures still being performed was 10, and the maximum reported savings after four weeks was 3535 kWh. In Table 12 you can furthermore see how many measures participants stated they started with (mean 1.45, *SD* 1.54), how many they still intended to do, how many they will likely not do, and corresponding saving metrics for each category. These two categories of 'intentions' and 'started', will not further be considered in the follow-up results section (only the items participants are 'still doing' (implying behaviors that were adopted, or actions that were completed). 30 participants did not complete any measures, of which 86% were non-homeowners (compared to 74% of the original and follow-up samples). We did not observe, for example, that this group indicated that they started with more measures than average, either.

Table 12

Choices in the follow-up study (N=160)

	Mean	*SD*	Median
Number of items still doing	2.07	1.92	2
Still doing kWh	315.8 kWh	538.2 kWh	95kWh
Number of items started	1.45	1.54	1
Started kWh	349.8 kWh	1080.3 kWh	50 kWh
Number of items intending to do	2.91	2.15	2
Intended kWh	1314.8 kWh	2852.8 kWh	340 kWh
Number of items not planning to do	1.81	2.20	1
Not doing kWh	1402.2 kWh	3582kWh	152.5 kWh
Number of previous items	8.23	4.18	8
Previous items kWh	3382.7 kWh	4751.8 kWh	1472.5 kWh

From Table 12, we see that on average, participants indicated that they still performed an item more often than indicating that they no longer planned on performing an item (2.07 against 1.81). However, the latter category has much higher average savings than the first one (as well as higher median savings), indicating that there might be a difference in what kind of measures people tend to give up on (e.g. with higher kWh savings).

We did not observe large differences in average item responses between the follow-up study and the initial study (As could be seen in Appendix E), indicating that the ones who participated in the follow-up study were likely not a subset of people who evaluated the system much more positively than average.

We repeated the same correlation matrix of personal characteristics, values, and system interactions as for the initial study, but now with actual savings, as can be seen below in Table 13.

Table 13

(Pairwise) Correlation matrix follow-up study (N=160)

Variables	(1) NEP	(2) IMS	(3) Male	(4) age	(5) kWh	(6) home	(7) en. lab	(8) # items	(9) amt	(10) Edu	(11) # current
(1) NEP score	1										
(2) IMS score	**-.23***	1									
(3) Male	**-.28***	.07	1								
(4) Age	.01	-.21*	.01	1							
(5) Actual savings	-.10	.08	.15	-.07	1						
(6) Homeowner	**-.20***	-.05	.12	.40*	.05	1					
(7) Energy Label	**-.22***	.01	.17*	.19*	**.20***	**.29***	1				
(8) Actual # of items	-.02	.14	.14	-.07	**.61***	.00	**.19***	1			
(9) Goal amount	.00	-.08	-.02	-.03	.11	-.02	.13	**.22***	1		
(10) Education	.11	-.07	-.06	-.14	-.01	-.08	-.05	.09	.00	1	
(11) Number of current items initial study (ability)	-.02	-.10	**.19***	**.21***	**.22***	.08	**.30***	**.17***	.06	-.08	1

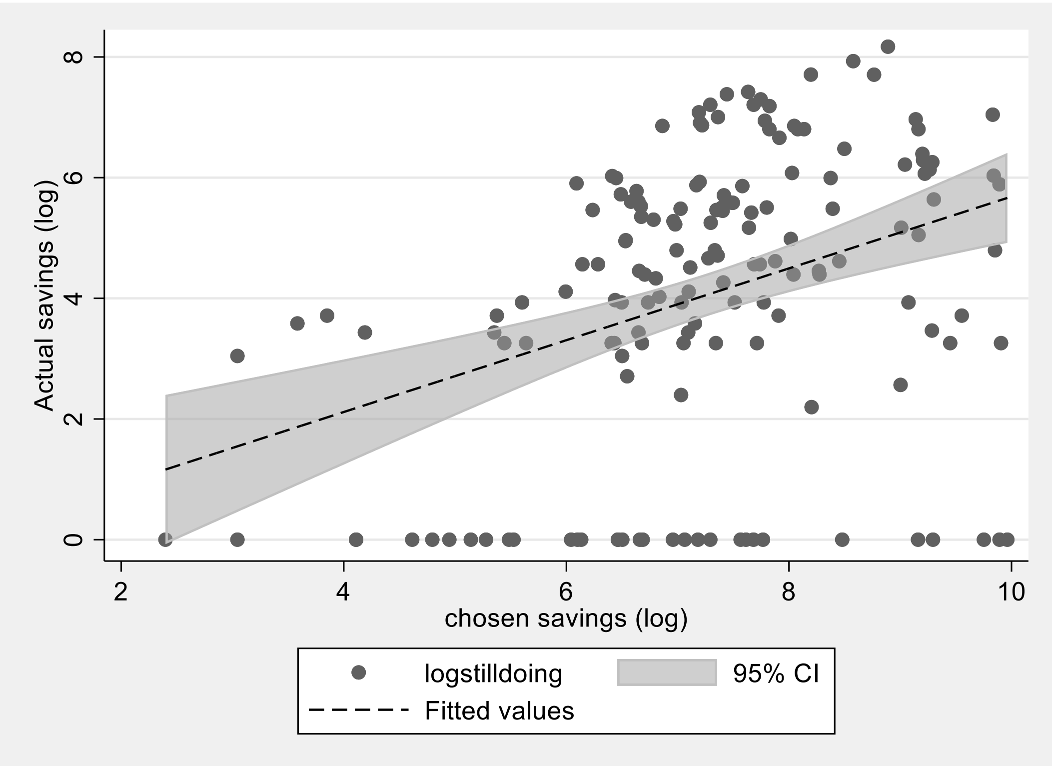
** shows significance at p<.05*

For the actual savings after four weeks, we found that a higher energy label is related to higher actual savings, and that higher ability is related to higher savings after 4 weeks. Gender, age, education level, NEP (pro-environmental values), or money importance values did not influence the extent to which participants performed energy-saving measures in the follow-up study.

When we compared the chosen savings from the initial study with the actual savings of the follow-up study, we did see a trend with higher chosen savings leading to higher actual savings (Figure 21).

Figure 21

Correlation between chosen savings and actual savings (not concerning those with chosen savings of 0 kWh).

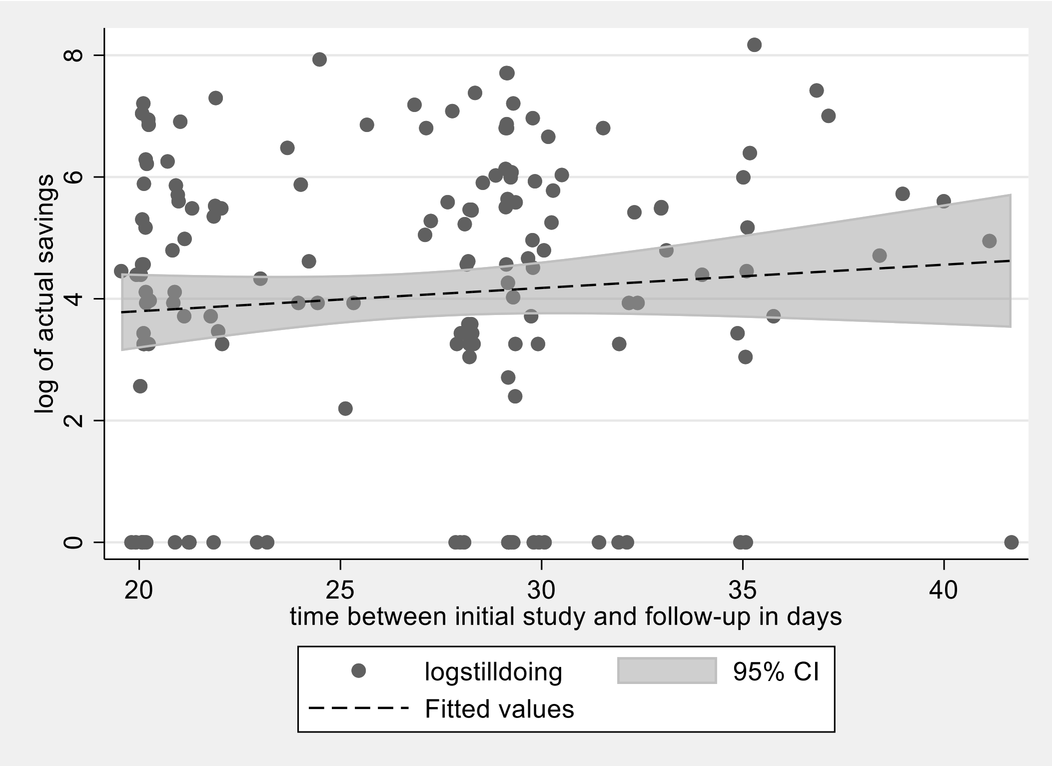


According to a pairwise correlation, this relationship was indeed significant with a coefficient of β = .36 and *p*<.0001 (However, without log transform, the relationship was not significant with a coefficient of .12 and p>.10)

Furthermore, there were approximately three to five weeks between the initial study and the follow-up, because the initial study ran over a longer time period than the follow-up (over around two weeks versus a timespan of around one week). We plotted the time between the initial study and the follow-up study per participant, against the eventual actual savings (Figure 22).

Figure 22

Actual savings over time between initial study and follow-up study



Although there might seem to be a slight upward trend here, where people perform more items with more time having passed, a pairwise correlation between those two variables was not significant (Coefficient=.09, p=.28).

4.2.2. Goal Conditions

We again compared the savings between goal conditions and goal amounts, as can be seen in Table 14. On average, people saved 341 kWh in the goal condition and 294 kWh in the no-goal condition. Averaged over all the goals, 10% of people already achieved their goal within these four weeks. We also see that with increasing goal height, people seem to perform an increasing number of measures.

Table 14

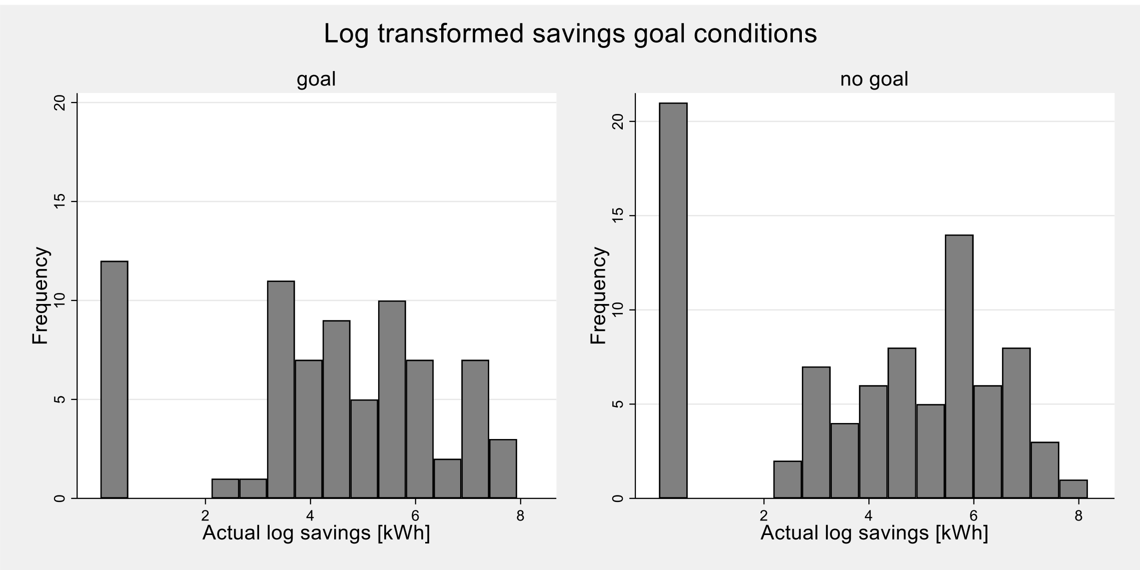
Savings and interactions per goal condition and goal amount in the follow-up study

Goal Amount	*Mean Savings (kWh)*	*SD*	Median savings (kWh)	*Mean # of items*	*SD*	*Median # of items*	*N*	*% goal reached*
No goal	294.0	(515.6)	85	1.79	(1.79)	1	85	
Goal	340.6	(565.3)	95	2.38	(2.02)	2	75	9.3%
Goal of 600	*299.7*	*(523.2)*	*62.5*	*1.97*	*(2.09)*	*1*	*34*	14.7%
Goal of 1200	*246.9*	*(432.7)*	*97.5*	*2.56*	*(1.88)*	*2*	*34*	2.9%
Goal of 1800	*993.9*	*(9.21)*	*1035*	*3.57*	*(2.07)*	*3*	*7*	14.3%
Total	315.8	(538.24)	95	2.07	(1.93)	2	160	

The saving-distributions (logarithmic) per goal condition can be seen in Figure 23. We observe two groups: one of people who performed no items, as discussed in the descriptive statistics section, and another of people who did perform items, which show a normal distribution.

Figure 23

Savings per goal condition



We repeated the rank sum test for savings between goal conditions, and once again found no difference in savings between the goal condition and the no-goal condition (*p* = .46). What we did observe is that in the no-goal condition, 20 out of 85 participants performed no items at all. Averaged over all goal conditions, only 10 out of 75 participants performed no items (in the graph, one sees a higher number of people without savings, because some participants ended up performing measures that had no net savings). This, and saving quartiles, can be seen in Table 15.

Table 15

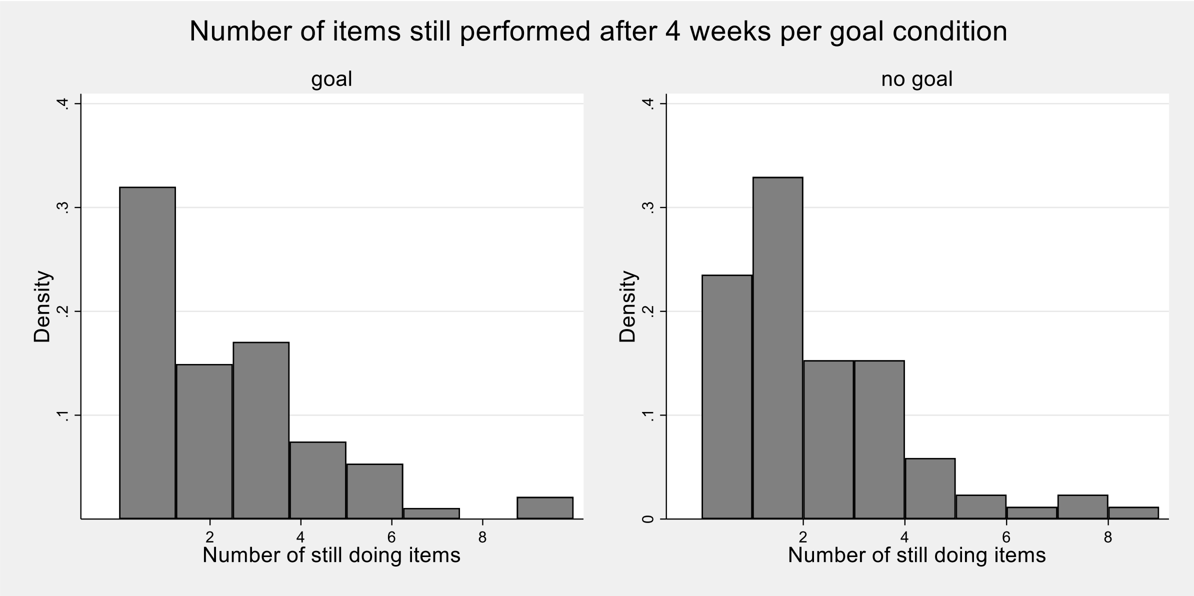
Quartiles per goal amount, and number of people per condition without savings.

Goal amount	p25 (Savings [kwh])	p25 (Savings [kwh])	p25 (Savings [kwh])	Number of people who achieved 0 kWh savings
No Goal	8	85	350	20 (out of 85)
600	25	62.5	322	6 (out of 34)
1200	35	97.5	235	3 (out of 34)
1800	225	1035	1320	1 (out of 7)
Total	25	95	352.5	30 (out of 160)

When we look at the distribution of the number of items in Figure 23, we observe a more gradual distribution of the number of items between goal conditions, no longer observing the distinct sub-group without savings that we observed in Figure 24. That strange distribution might therefore have been caused by variance in savings of items themselves, rather than by participant characteristics.

Figure 24

The number of items still performed after four weeks, per goal condition.



4.2.3. Signpost Conditions

We furthermore compared savings and interactions per signpost condition as seen in Table 16. In the Euro signpost condition, average savings appear to be lower, while savings appear to be highest in the CO2 condition. This was reversed in the initial study, where the CO2 signpost had the lowest average savings.

Table 16

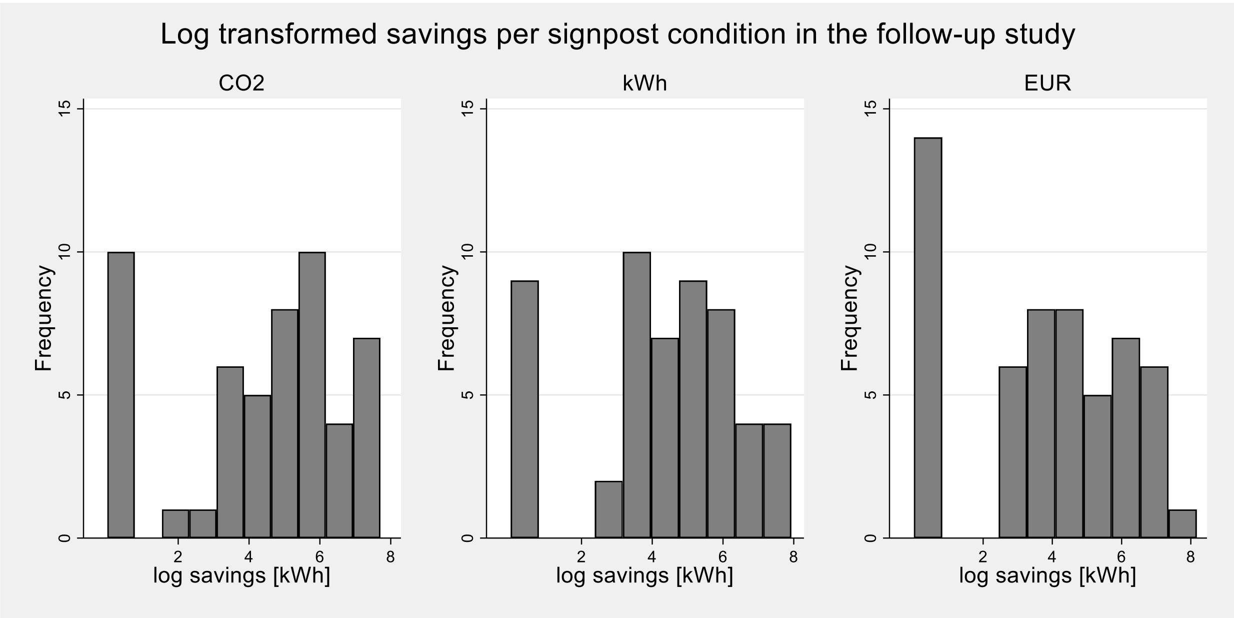
Savings and interactions for each signpost condition in the follow-up study.

signpost	Mean Savings (kWh)	*SD*	Median	# of items	*SD*	Median	Freq.
CO2	347.7	507.8	141	2.17	2.08	2	52
EUR	268.2	538.3	52	1.78	1.65	1	55
kWh	334	573.9	95	2.26	2.02	2	53
Total	315.8	538.24	95	2.07	1.92	2	*160*

From the graphs in Figure 25, we see that, although the main distributions appear similar, there were more people in the Euro condition who ended up not performing any items at all. To see if there might be a bias in the follow-up data that caused this strange distribution, we checked several demographic differences between conditions. The mean ages in all conditions were similar, (28.5 for the kWh condition, 30.4 for the Euro condition, and 30.3 for the CO_2 condition with *SD*s between 12 and 14), as were gender ratios, but we did see that the Euro condition had relatively little homeowners (18% compared to 32% in the kWh condition and 25% in the CO_2 condition), which might have influenced results somewhat.

Figure 25

Savings per signpost condition in the follow-up study.

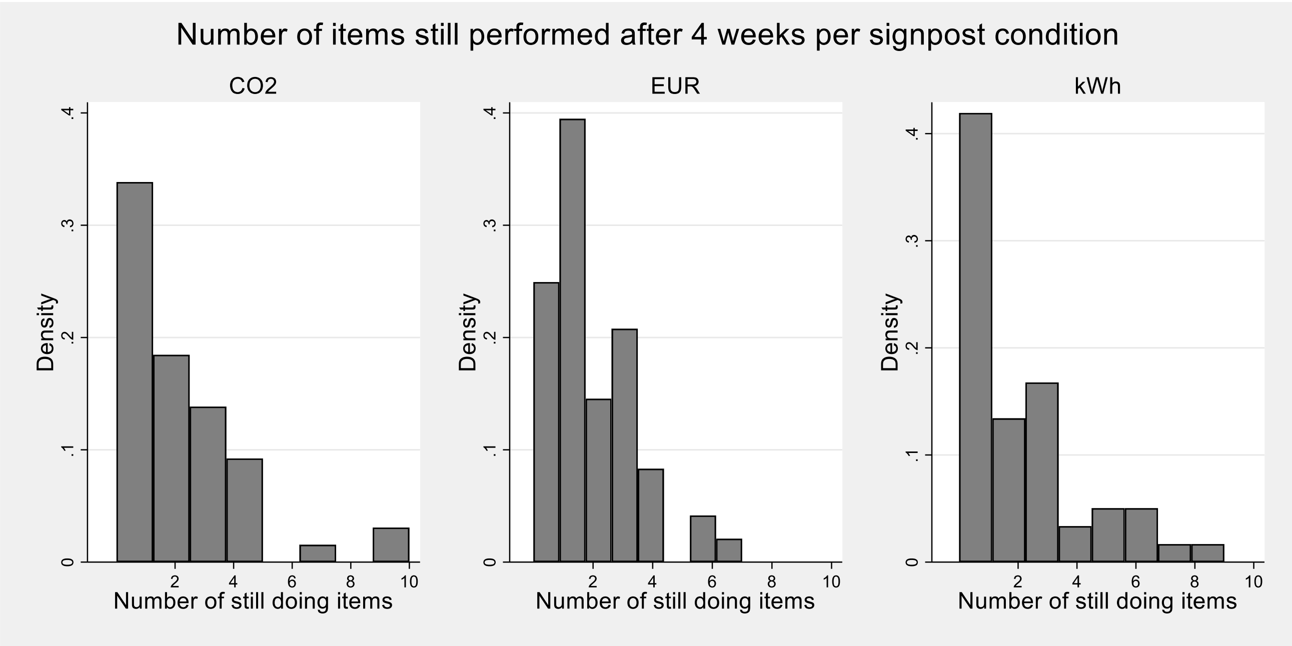


We repeated the ANOVA between signposts, after a quadratic transformation, and the overall model was not significant ($F(2,157)=0.77$, $p=.46$), meaning that the apparent lower savings in the Euro condition, and higher savings in the CO_2 condition, were not the result of a significant effect, whereas the previously observed difference in the initial study (with lower savings for the CO_2 signpost) was significant. Therefore, the influence of signposts by themselves, was only limited to the chosen savings, and not on actual savings (although we did not have an hypothesis on signposting effects without value interactions).

In the graph below in Figure 26, you can furthermore see the number of items performed after four weeks in each signpost condition, which, similarly as with the goal setting conditions, looks more gradual than the savings per signpost condition.

Figure 26

The number of items still performed after four weeks, per signpost condition.



4.2.4. Direct Effects from Signposts and Goals on Actual Savings

We repeated the regression with interaction effects from section 4.1.6 (Table 11), using actual savings as the dependent variable, as seen in Table 17.

Table 17

Robust linear regression predicting the actual savings based on experimental conditions and NEP and IMS scores.

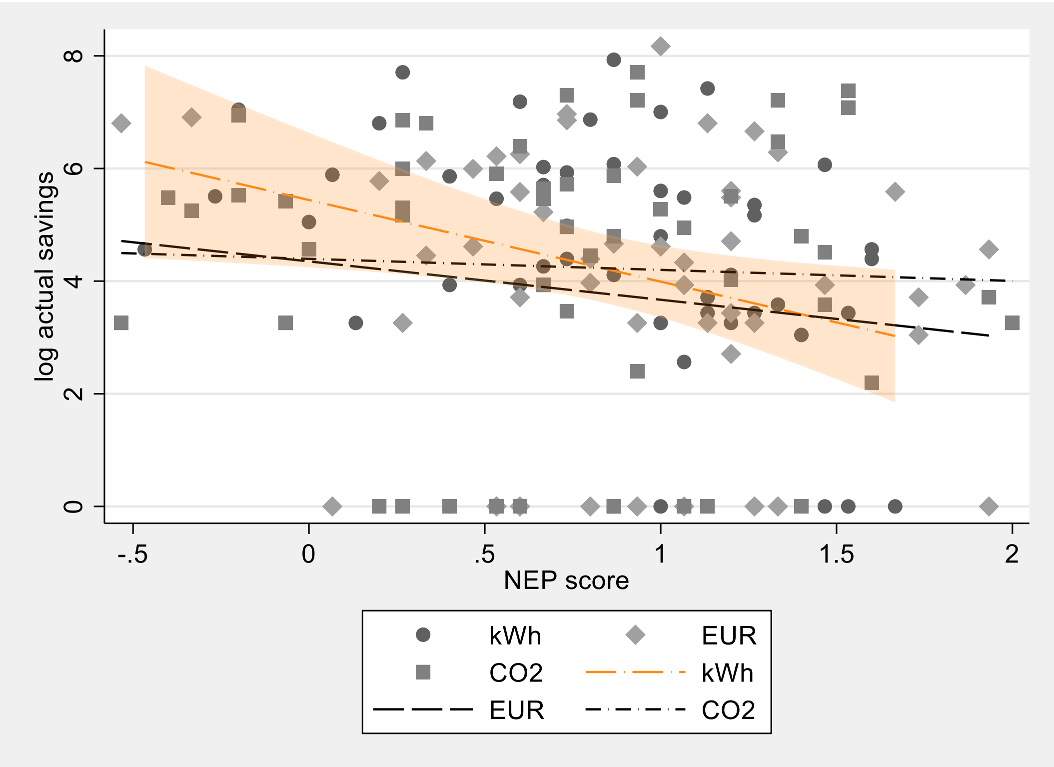
Variables	Log Actual savings. Coef.	β	*SE*	95% CI
IMS score	0.46	0.14	(0.50)	[-0.53; 1.45]
Signpost				
CO2	-1.30	-0.25	(1.00)	[-3.27; 0.67]
EUR	-0.27	-0.05	(1.00)	[-2.25; 1.72]
Signpost # IMS score				
CO2	-0.20	-0.04	(0.67)	[-1.51; 1.12]
EUR	-0.49	-0.09	(0.72)	[-1.91; 0.94]
NEP score	-1.46*	-0.33*	(0.57)	[-2.59; -0.33]
Signpost # NEP score				
CO2	1.33	0.27	(0.74)	[-0.13; 2.80]
EUR	0.78	0.17	(0.82)	[-0.85; 2.40]
Goal condition				
No goal	-0.24	-0.05	(0.66)	[-1.54; 1.06]
Goal Condition#signpost				
No goal # CO2	0.41	0.06	(0.95)	[-1.48; 2.29]
No goal # EUR	-1.10	-0.19	(0.94)	[-2.96; 0.76]
Constant	5.46**		(0.76)	[3.95; 6.96]
Observations	160			
R-squared	0.081 (Prob>F=0.08)			

*** p<0.01, *p<0.05*

In this regression, the kWh and goal conditions are the baseline. We find that for this baseline, an increasing NEP score results in decreased savings (β = -.33, *p*<0.05). The interaction effect between the CO2 signpost and the NEP score, is no longer significant (which was already a very small effect in the initial study (β = .10, *p*<.01), and was only observed in the SEM model and not in the regression). The remaining, non-significant relations between signposts and nep scores on actual savings can be seen in Figure 27. We still observe a downward slope for the kWh condition, seemingly stronger than in the initial study (Figure 20), although the 95% confidence interval is very wide and overlaps with the fit lines of the other two signposts. The downward slope for the CO2 condition seems to have become less steep than before.

Figure 27

Scatter plot of NEP score and log of actual savings, per signpost condition



We did not repeat the regression with self-efficacy and choice satisfaction as dependent variables, as this would concern the same regression as before on a reduced sample (participants did not answer these questions a second time for the follow-up study).

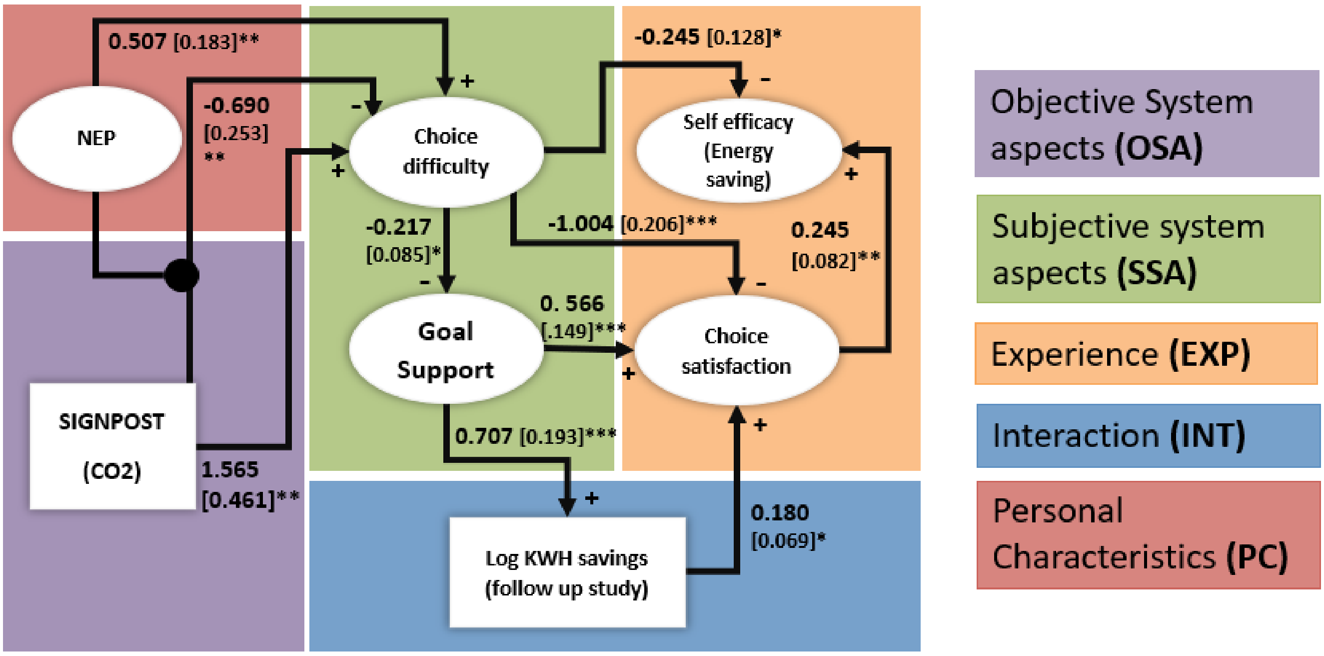
4.2.5. Factor Analysis and Structural Equation Model (SEM)

We furthermore repeated the confirmatory factor analysis with mPlus, the results of which could be seen in Table 3 in the materials section (3.4). The items that did not have factor loadings in the initial study, were not considered, to keep constructs the same for both studies. The results were similar to the initial study as the same scale answers were used, albeit only for the people who returned for the follow-up study. The full outputs of the CFA/SEM can be found in Appendix J (Please note that the choice difficulty scale is reversed in this output).

We repeated the SEM analysis for the follow-up data, but now with the actual savings after four weeks, of which the results can be seen in Figure 28. For the full mPlus output, refer to Appendix J.

Figure 28

Structural equation model (SEM) for the follow-up study. Coefficients are depicted as numbers on the arrows, and standard errors in brackets. (Design in line with the user-centric evaluation framework by Knijnenburg and Willemsen (2015).



*N=160, *** p < 0.001, ** p < 0.01, * p < 0.05*

The fit of this SEM model was good, with χ^2 (*N*=160) of 252.1 (*p*<.01), RMSE of .044 (90% *CI* between .027 and .058), comparative fit index CFI of .971 (>.96) and Tucker-Lewis Index TLI of .967 (>.95).

As most of these factors are largely the same as in the initial study (and based on the same answers from the initial study), we will focus on the indirect effects of signposts and interactions on actual savings in the follow-up study. As with the initial study, we found no effects of goal conditions on any of the variables, thus not finding support for hypotheses 1 until 6.

Compared to the initial study, there is no longer an effect of the Euro signpost on choice difficulty (p>.30) or an interaction effect between the NEP score and the Euro signpost on choice difficulty (p=.20). Direct effects between choice difficulty, goal support, choice satisfaction, and self-efficacy are still present and very similar to the initial study.

Actual Savings After 4 Weeks. We see no direct effects of signposts, NEP scores, or interactions on actual savings. The indirect effect of the interaction Euro and NEP score on savings is no longer significant (p>.25), and the indirect of the CO2 signpost and NEP score on savings through choice difficulty and goal support, is also no longer significant, albeit close to (β = .106, p=.089). The same is true for the indirect negative effect of the NEP score on savings, via choice difficulty and goal support (β = -.078, p=.080), and the indirect effect of the CO2 signpost on savings via choice difficulty and goal support (β = -.24, p=.070). This effect would have been almost equal in size as in the initial study. The indirect effect of the Euro signpost on savings is no longer significant (.425). This effect was already quite weak in the initial study (β = .18) and was exactly significant there at p=.05, so this is not entirely surprising.

To see how sample size affects the significance of effects in this follow-up study, we can look at self-efficacy. We do still see an indirect interaction effect of the CO2 signpost with NEP score on self-efficacy (β = .36, p<.05), which is mostly explained by the pathway via (decreased) choice difficulty and (increased) choice satisfaction (β = .36, p<.05). The pathway through choice difficulty is however no longer significant (β = .38, p=.087). As these factors were all based on the same answers as the initial study (just from a reduced sample), the likely reason that the latter pathway is no longer significant, is the reduced power of this study. This might be the same for previously discussed effects. However, corrections for multiple tests might have been needed even if we had observed significant interaction effects on actual savings, due to performing multiple tests across two studies. Therefore, we do not want to interpret this close-to-significant finding between CO2 and NEP on actual savings, as support for hypothesis 7.

5. Conclusion and Discussion

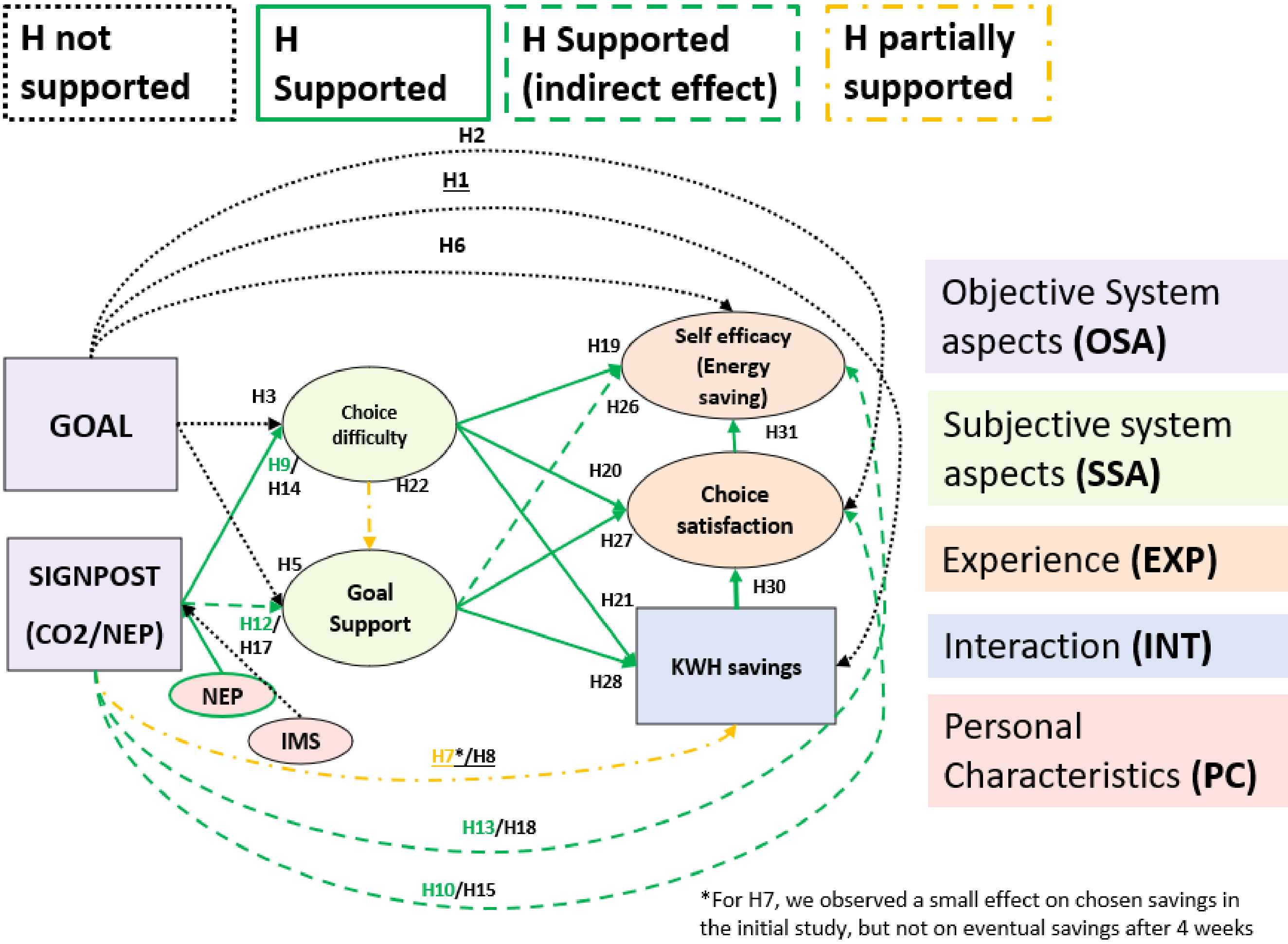
In this study, we tested a Rasch-based energy-saving recommender system, and various manipulations of this system aimed at increasing user satisfaction and decreasing household energy usage of participants. We tried to answer the question to what extent signposts and goal setting can influence user behavior and user experience in such a system, and how these effects can be explained by how users perceive this system. For this, we asked participants to choose energy-saving measures and evaluate the system, and after four weeks we asked which of these measures they ended up performing.

5.1. General Discussion

We found that the system worked quite well in terms of helping users save energy, with an average of 316 kWh saved after 4 weeks (and a median of 95 kWh), and in general, users seemed to appreciate the system based on their responses to user-evaluation statements (Appendix E). While we did find support for several hypotheses, we found no effects of goal setting on any of the outcome variables, and the effects of signposting were not entirely in line with our expectations, either. For example, we found various effects of signposts in isolation (such as that people chose higher savings with the Euro signpost), whereas in our hypotheses we only expected signposts to have an effect through interactions with user values. Because there were quite a few hypotheses, you can find the ones that were supported in Figure 29 below in green, and the ones that were not supported in (dotted) black. The system satisfaction scale and perceived feasibility scales were omitted, as they were not included in the analyses due to high correlations. The green lines from the 'signpost' hypotheses refer to interactions between the CO2 signpost and NEP scores.

Figure 29

Conceptual model with study outcomes.



5.2. Discussion of the Theoretical Framework

We observed various indirect effects of system characteristics on subjective perceptions of the system, and in turn, we saw that these subjective perceptions influenced experience and behavior outcomes. We observed that decreased choice difficulty led to increased goal support, and that increased goal support led to higher savings and higher choice satisfaction. Decreased choice difficulty furthermore led to increased choice satisfaction; in fact, these measures were so strongly correlated that we cannot confidently say they measured two distinct things. Both increased choice satisfaction, and decreased choice difficulty, led to higher energy-saving self-efficacy. We also saw that increased savings led to increased choice satisfaction, which was also observed in the study by Knijnenburg et al. (2014).

What this all shows, is that the eventual outcomes of the system, can be explained by how the system impacts the subjective experience of participants, and how this subjective experience then influences outcomes. That all effects from system aspects on experience outcomes, were mediated by subjective system aspects, is also in line with the findings of the Knijnenburg et al. (2014) study. While signposts eventually did not seem to influence savings, they did influence self-efficacy and choice satisfaction (though, these were measured in the first study, so we cannot be sure that this effect remained after four weeks), and these effects were entirely mediated by changes in choice difficulty and goal support (of which goal support is similar to system satisfaction in the Knijnenburg et al. (2014) paper). This also shows that there were likely no other constructs that mediated these relationships; in which case, we would have found direct effects that could not be explained by system characteristics.

5.3. Discussion of the Effect of Goal Setting

Our first research question concerned the effect of guided goal setting on choice satisfaction, energy-saving self-efficacy, and energy savings in a Rasch-based energy recommender system. We had expected that guided goal setting would on the one hand motivate people to achieve higher savings, while the goal suggestions would ensure that these goals would be reachable. We, therefore, expected that experience outcomes (choice satisfaction and energy-saving self-efficacy) would also improve, and that all these effects would be mediated by subjective system aspects (decreased choice difficulty and increased goal support).

However, we found no effect of goal setting on savings, choice satisfaction, environmental self-efficacy, or any of the mediating subjective system aspects. An explanation of this might lie in the height of the goals that participants could choose from. When we compare the savings of this study to that of the study by Bams in 2018, we see that chosen savings in our initial study were much higher than before; 3057 kWh with a median of 1350 kWh now, compared to around 1200 kWh with a median of around 250 kWh (or 660 kWh for those who chose at least some amount of savings), in 2018. It might therefore be that the goals were simply too low and did not encourage higher savings than what participants otherwise would have selected. A side note here, is that in all conditions, the majority of chosen savings were not completely realized after four weeks. Still, four weeks is a rather short time for making adaptations to one's house, for example, and people often indicated they still planned to do the majority of items or that they had made a start with them.

Despite not finding effects in the initial study, we had hoped that the participants in the goal-setting condition would have set more realistic goals for their situation, and therefore, that we might have found an effect in the follow-up study. This was not the case: after four weeks, we still did not observe a difference between the two goal conditions. While it could certainly be the case that goal-setting in this context simply is not effective, it might also have been partly due to the design of our system. Goal setting is often used in a more interactive system, and additionally, goal setting requires feedback, according to the goal-setting theory by Locke and Latham (2002), and previous research by Kim and Hamner (1976). This was quite sparse within our system. Participants got an email with a link to their chosen measures approximately one week after participation, but this is something very different from an app that continuously gives performance feedback. Furthermore, the goal was no longer visible in this review: it was a mere list of chosen items, with images, signposts, and descriptions, but it was not an interactive system.

5.4. Discussion of the Effect of Signposting

Our second research question concerned the effect of signposting on choice satisfaction, self-efficacy, and energy savings in a Rasch-based energy-saving recommender system. In our study, we compared the effect of three different signposts (kWh, CO2, and Euro) on experience outcomes and chosen savings in the initial study and actual savings in the follow-up study. We will first discuss the effect of signposts on savings, and then the effect of signposts on experience outcomes (self-efficacy and choice satisfaction).

Savings. In the initial study, we found that the CO2 signpost led to a lower amount of chosen savings than the kWh signpost, regardless of the value orientation of participants, albeit this was a small effect of $\beta = .18$. However, we did not observe such an effect for the actual savings after four weeks. This might be because the sample was smaller, and thus it was harder to observe an effect, or it might be that the kWh signpost motivated people (to a small extent) to choose more measures, but this motivation did not translate into motivation to actually perform these measures.

The effect that the CO2 signpost led to lower savings than the kWh signpost, was entirely mediated by how participants perceived the system. In the follow-up study, however, the average savings between signpost conditions did not significantly differ anymore. Perhaps, over time, the actions themselves become more salient whereas the signposts and possible value activations might have gone more into the background. A system that is more often used might show different results and might be able to maintain the effects seen in the original study.

Additionally, the translation of attributes automatically means that numbers change as well, and these might make it easier or more difficult to compare measures: kWh ratings had a seemingly wider range (from 0 to 8000 kWh) than CO2 ratings (from 0 to 2000 kg CO2) or monetary ratings (€0,- to €1200,-). Because of this, items with lower savings might have been easier to comprehend in the kWh condition than in the CO2 condition, causing these effects of signposts in isolation.

Interaction Effects of Signposts and Values on Savings. While we did observe these unexpected differences between the signpost conditions in terms of savings, the interaction effect with personal values was minimal. We did not observe substantial differences in savings based on NEP scores and signposts, except for a very small indirect effect ($\beta = .10$) in the initial study for the interaction of the NEP score and the CO2 signpost on savings. This effect was no longer significant in the follow-up study. The idea behind signposting was, that a certain translation of attributes activates certain values (Ungemach et al., 2018), after which people are expected to act more in line with those values. We therefore expected, for example, that people who scored higher on the importance of money scale, would obtain higher savings in the monetary condition than those who scored lower on the importance of money scale. The same effect was expected for those with stronger pro-environmental values, as measured by the NEP scale; that they would obtain higher savings in the CO2 condition as compared to people with weaker pro-environmental values.

There might be various reasons why we did not observe an interaction effect between signposts and values on actual savings after four weeks, and why the effect we observed between the CO2 signpost and NEP score on chosen savings in the initial study, was very small. What might be at play here is that people could have other, conflicting values that we did not inquire about. In addition to values, constraints on behavior might be at play, especially in our relatively young sample. No matter how concerned one is about the environment, large changes to one's dwelling such as installing solar panels or improving wall insulation, are simply not an option for many who hire accommodation. Furthermore, stronger pro-environmental values might not always translate into greater knowledge: someone might have a better grasp of the meaning of, say, 30kWh, than they do of '3kg CO_2', despite their concern with the environment. Especially since energy providers supply us with the kWh and gas usage metrics, rather than the amount of carbon dioxide we have emitted into the atmosphere, the CO_2 signpost condition might therefore not show a strong interaction effect with environmental values.

Lastly, many of the hypotheses were based on the assumption that those with a higher NEP score, would achieve higher savings in the CO_2 condition because this condition would supposedly be in line with their values. However, there is another underlying assumption here: that activating those values would somehow mean activating a desire to save more energy (and thus, that this 'motivation' would be activated in the CO_2 signpost condition). Based on these assumptions, one would also expect that those with higher NEP scores would choose higher goal amounts in the CO_2 condition. Something that we did not observe, either. It might therefore be that pro-environmental values are not always equivalent to a desire to save energy.

We did not observe any interaction effects from the IMS score and Euro signpost on savings, or on any other variables for that matter. When we look at the questions in the IMS scale, they seem to be more about the extent to which one values 'having' money, and not necessarily the extent to which one values 'saving' money. These might be two different things: someone might be very motivated to earn money, such that they can spend it on things they enjoy or not have to worry about their energy bill for example. Those with higher IMS scores, might therefore not always want to 'save money'. It might also be for example, that someone who scores high on this scale, might be more financially secure already if they have prioritized their financial well-being, and thus might have less immediate need to reduce their spending at this moment. Because we did not ask for income or financial position of participants, as this was considered sensitive data (and not directly related to our research questions), we were not able to control for these factors.

Signposts and User Experience. We also found isolated effects from signposts on experience outcomes, with the CO2 and Euro signposts leading to lower energy-saving self-efficacy as compared to the kWh signpost. This effect was mainly explained by an increased choice difficulty for these conditions and in part, because that contributed to decreased choice satisfaction. While we had not hypothesized these effects, it might be that the novelty of these signposts has played a part in this: People might be more used to making decisions on the basis of kWh usage of appliances, as opposed to CO2 or Euro usage metrics. This could explain why they experience more difficulty in these conditions as it concerns a more unfamiliar context. This might be an effect that would resolve with longer use of the system, as people would get more used to these signposts, but it might also be a general trend.

Interaction of NEP and CO2 on User Experience. For the experience outcomes, we did observe interaction effects with the CO2 signpost and NEP score, as we had hypothesized. In the CO2 condition, as compared to the kWh condition, we saw that the NEP score was positively correlated to self-efficacy, and we found the same for choice satisfaction; for this signpost, increasing NEP score led to increased choice satisfaction. While this would strictly provide support for our hypotheses, we observed that these effects seemed to be mostly due to negative correlations of NEP and experience outcomes within the kWh condition. Because we used this kWh condition as a baseline, in comparison, there seemed to be effects in the CO2 condition. However, this was not the case: Within the CO2 condition, increasing NEP score did not lead to better experience outcomes. Thus, the idea that the CO2 signpost would activate user values after which users would act in line with these values, is difficult to support from the data we observed. However, for practical implications, these effects might still be relevant, despite that the underlying mechanisms might have been different than expected.

NEP Score and User Experience. Overall, we saw that an increasing NEP score led to decreased experience outcomes (decreased self-efficacy and decreased choice satisfaction). While we did not have explicit expectations for the NEP score by itself, we did not expect an increasing NEP score to result in worse experience outcomes. It might be that those with stronger pro-environmental values, have higher expectations of such a system, or are more critical of this system. It might also be that they are less satisfied with their chosen measures, precisely because they want to do more and they might feel that they cannot live up to these goals.

5.5. Limitations of the Study

The study had several limitations, which might have impacted the extent to which this study could answer the relevant research questions. There was a biased sample, and the study relied entirely on self-reporting. Additionally, we think that a portion of the participants might have misunderstood the experiment.

5.5.1. Large Variance in Items

Similarly to previous Rasch-based recommender studies, we did find differences in averages between conditions, e.g. people in the goal-setting condition achieved on average 340 kWh in savings after four weeks, compared to 294 kWh in the no-goal condition. However, these were not significant due to the large variance, and large standard deviations (+- 500 kWh), in the data. This might be due to the nature of the system: there is a great variety of measures with very low and very high savings (from 0 to 8000 kWh, with an average of 516 kWh and *SD* of 1333 kWh), possibly resulting in the fact that there is also great variance in the savings chosen by participants. We had hoped that in the goal condition, there would be less variance from guiding people towards a certain saving amount, but this was not the case.

5.5.2. Biased Sample

Despite using both the Prolific and Archie participant databases, we obtained a relatively young sample, with a mean age of 30.5, and of which 75% was below 32 years old. Most of our participants were non-homeowners (77%). Therefore, several efficiency items might not have been applicable for everyone as they required modifications to a dwelling, and this would usually not be possible for someone who rents a living space. This was also something that several participants stated in the open feedback form. From the data, we could see that several curtailment measures were more popular than in the previous Rasch energy recommender studies by Starke (2017-2021) and Bams (2018) and that certain efficiency measures were less popular than before (as seen in Figure 11). For this previous study, the average age was 34.2 for the part that determined item difficulties, and 54.8 for the recommender system part (Bams, 2018). The whole application therefore might represent a less realistic scenario than it would to an older sample, possibly influencing the deliberation with which participants choose measures, and thereby possibly reducing a possible effect of system manipulations.

Our study furthermore relies entirely on self-reporting, which might bias results somewhat towards reporting higher savings than actual. It might have been, that those who chose a higher amount of savings in the initial study, were more inclined to also report having performed more measures in the follow-up study to appear consistent in what they said they would do, and what they actually did. However, without obtaining actual energy usage data, we cannot really know if this was the case. Note that in the follow-up, participants were not reminded of their self-chosen goals, nor did we highlight how their actual performed measures contributed to these goals or to a total saving amount. Therefore, we do not expect that these participants focused on 'reaching' their goals in the follow-up study. We also saw that only a fraction of participants in fact reached their goals.

5.5.3. Instructions

It also seemed as if several participants might have misunderstood the assignment in the phase that asked them to select new actions. Several participants selected 20 actions, and the overall sample had an average of 7.5 actions. This average was much higher than in the previous study by Bams in 2018. It might have been that participants thought that they had to choose between *'I already do this'* and *'I will do this'* and did not perceive the list as one of recommendations, of which to choose amongst. In fact, 33 participants clicked on every single item (either to select it, or to remove it). Given that the average savings were also much higher, it might have been useful to instruct participants to only choose measures they genuinely planned on performing. In the goal condition, participants might have better understood that they were not expected to click every single item (as they would have likely reached their goal before this point), whereas, in the no-goal condition, participants did not have such a focal point. This might explain why we saw (although not significant) higher average chosen savings for the no-goal condition than for the goal condition.

5.5.4. Insufficient Power for the Follow-Up

To properly fit the SEM model, we needed 210 participants. However, the follow-up response rate was around 80%, leaving us with 160 participants after outlier removal, and insufficient power for this new analysis. Although the analysis 'ran' (thus, it could be fitted), we saw several effects that were close to 0.05 but slightly larger. Because of the 0.05 cut-off value, these effects cannot provide support for the hypotheses we had. We do not know if this was due to the effects not being there, or due to insufficient power to show significant effects.

5.6. Practical Implications

The current study does not find support for the idea that goal setting would be effective in the current system. However, we found no effects, thus, also no negative effects. Given the large body of previous research on goal setting, it might still be interesting to experiment with the height of the goals and gather more feedback for a possible more effective implementation. It should be considered, however, that no goal might be better than a goal not reached, in terms of system satisfaction.

For signposting, besides that the effects were due to differences for the kWh signpost, we recommend that systems consider user value orientations when choosing signposts. While the CO2 and EURO signposts seemed to work equally well for everyone, the kWh signpost seemed to work especially well for people lower on the NEP scale, in terms of self-efficacy, reduced choice difficulty, and increased choice satisfaction. Therefore, it would be helpful to consider the target population of a system and adapt the system accordingly. A side note here is that the differences between signposts are most distinct for the lower end of the NEP scale, whereas on the higher end of the NEP scale, these (opposite) effects, are minimal, and might only affect those with very high NEP scores. Overall, the kWh signpost seems to be best for most users in terms of experience outcomes, despite this not being our initial research focus.

5.7. Future Research

In the future, it might be interesting to look at system satisfaction after four weeks, alongside actual savings. This is because satisfaction might have changed, depending on the extent to which people have fulfilled their initial plans and reached their self-chosen goals. Increased system satisfaction for either condition might mean that, despite observing no immediate effects, one of the groups could be inclined to use the system for a longer period, and this could eventually still impact energy savings. In addition to that, it would be helpful to start with a larger than necessary sample, to ensure sufficient responses in the follow-up study (assuming a response rate of around 80%).

A larger sample size could furthermore be useful to see if the differences in means that we observed could translate into significant effects, despite large variances in the data. However, in the context of a formal study with participant recruitment, this might not be feasible in terms of time and money needed. In the context of an existing website where A/B testing is performed, it might be.

Furthermore, a system that operates more like an app that provides participants with continuous feedback and allows participants better insight into their own achievements, might see a greater effect of goal setting than our one-off experiment with a one-month follow-up.

Additionally, a replication study that compares the effect of signposts by themselves, in the context of such an app-based system, might provide more insight into the differences observed in the initial study, as we did not have any formal hypotheses on this. It might be that besides value activations, other effects are at play that result in certain signposts being more effective for most users.

Lastly, in line with what several participants requested in the feedback form, it would be interesting to see if effects can be observed in a system that considers more aspects of a participant's situation, like dwelling type, rental status, existing energy label, etcetera. This would mean that the system would no longer be based on a purely one-dimensional construct. The Rasch scale could still be used, with the simple addition of filtering out some irrelevant items. Controlling for dwelling type in the analyses, might then also be necessary. A shortcut to this approach would be, to only let homeowners participate, who would generally have more possibilities in terms of energy saving. A more qualitative approach might help to understand which other adaptations to the system users might appreciate. Especially since this system is very distinct from smart-meter-based recommenders, starting from such a viewpoint might be more effective in identifying possible areas of improvement than starting from a theoretical perspective. Not only could eventual system adaptations then improve user satisfaction; providing users with measures they are in fact able to perform, might also result in stronger effects of system manipulations and more informative data from decisions made in a more realistic setting.

6.References

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2007). The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. *Journal of Environmental Psychology*, *27(4),* 265–276. https://doi.org/10.1016/j.jenvp.2007.08.002

Bams, L. P. (2018) *Exploring the determinants of energy-saving behavior to nudge users of a Rasch-based energy-saving recommender to higher energy savings* (Master's thesis, Eindhoven University of Technology, Eindhoven, Netherlands). Retrieved from https://pure.tue.nl/ws/portalfiles/portal/107655057/Master_s_Thesis_Luc_Bams_0804795.pdf

Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-efficacy: The exercise of control. Retrieved from https://www.academia.edu/28274869/Albert_Bandura_Self_Efficacy_The_Exercise_of_Control_W_H_Freeman_and_Co_1997_pdf

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology, 41(3),* 586–598. https://doi.org/10.1037/0022-3514.41.3.586

Boudet, H. S., Flora, J. A., & Armel, K. C. (2016). Clustering household energy-saving behaviours by behavioural attribute. *Energy Policy, 92,* 444–454. https://doi.org/10.1016/j.enpol.2016.02.033

Brandsma, J. S., & Blasch, J. E. (2019). One for all? – The impact of different types of energy feedback and goal setting on individuals' motivation to conserve electricity. *Energy Policy,* *135*, 110992. https://doi.org/10.1016/j.enpol.2019.110992

Eccles, J. S., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, *53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Elliot, Andrew J., and Harackiewicz, J. M. (1994) "Goal setting, achievement orientation, and intrinsic motivation: a mediational analysis." *Journal of personality and social psychology 66*, no. 5 968. https://doi.org/10.1037/0022-3514.66.5.968

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a New General Self-Efficacy Scale. *Organizational Research Methods, 4(1),* 62–83. https://doi.org/10.1177/109442810141004

Cervone, D., Jiwani, N., & Wood, R. (1991). Goal setting and the differential influence of self-regulatory processes on complex decision-making performance. *Journal of personality and social psychology, 61(2),* 257. https://doi.org/10.1037//0022-3514.61.2.257

Domalewska, D. (2021). A longitudinal analysis of the creation of environmental identity and attitudes towards energy sustainability using the framework of identity theory and big data analysis. *Energies, 14(3),* 647. https://doi.org/10.3390/en14030647

Dosio, A., Mentaschi, L., Fischer, E. M., & Wyser, K. (2018). Extreme heat waves under 1.5 °C and 2 °C global warming. *Environmental Research Letters*, *13(5),* 054006. https://doi.org/10.1088/1748-9326/aab827

Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Jones, R. E. (2000). New Trends in Measuring Environmental Attitudes: Measuring Endorsement of the New Ecological Paradigm: *A Revised NEP Scale. Journal of Social Issues, 56(3),* 425–442. https://doi.org/10.1111/0022-4537.00176

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. Annual *Review of Psychology, 53(1),* 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Fraser, A. (2023). Success, failure, and information: how households respond to energy conservation goals. *Journal of the Association of Environmental and Resource Economists*, *10*(1), 95-120. https://doi.org/10.1086/721094 Retrieved from: https://www.alastairfraser.com/research

Franzen, A., & Mader, S. (2022). The Importance of Money Scale (IMS): A new instrument to measure the importance of material well-being. *Personality and Individual Differences, 184,* 111172. https://doi.org/10.1016/j.paid.2021.111172

Harding, M., & Hsiaw, A. (2014). Goal setting and energy conservation. *Journal of Economic Behavior & Organization,* 107, 209–227. https://doi.org/10.1016/j.jebo.2014.04.012

Himeur, Y., Alsalemi, A., Al-Kababji, A., Bensaali, F., Amira, A., Sardianos, C., ... & Varlamis, I. (2021). A survey of recommender systems for energy efficiency in buildings: Principles, challenges and prospects. *Information Fusion*, *72*, 1-21. https://doi.org/10.1016/j.inffus.2021.02.002

Intergovernmental Panel On Climate Change. (2021, August 9). Climate change widespread, rapid, and intensifying – IPCC. https://www.ipcc.ch/2021/08/09/ar6-wg1-20210809-pr/

International Energy Association IEA. (2020). Final consumption – Key World Energy Statistics 2020 – Analysis - IEA. IEA. https://www.iea.org/reports/key-world-energy-statistics-2020/final-consumption

International Energy Association IEA. (2022). Methane and climate change – Global Methane Tracker 2022 – Analysis - IEA. IEA. https://www.iea.org/reports/global-methane-tracker-2022/methane-and-climate-change

Karlin, B., Davis, N., Sanguinetti, A., Gamble, K., Kirkby, D., & Stokols, D. (2014). Dimensions of Conservation: Exploring Differences Among Energy Behaviors. *Environment and Behavior*, *46(4),* 423–452. https://doi.org/10.1177/0013916512467532

Kim, J. S., & Hamner, W. C. (1976). Effect of performance feedback and goal setting on productivity and satisfaction in an organizational setting. *Journal of Applied Psychology, 61(1),* 48. https://doi.org/10.1037/0021-9010.61.1.48

Knijnenburg, B. P. (2009). Adaptive advice: adapting a recommender system for energy-saving behaviors to personal differences in decision-making. [Master thesis, Eindhoven University of Technology]

Knijnenburg, B. P., Reijmer, N. J. M., & Willemsen, M. C. (2011). Each to his own: How different users call for different interaction methods in recommender systems. *Proceedings of the Fifth ACM Conference on Recommender Systems, 141–148*. https://doi.org/10.1145/2043932.2043960

Knijnenburg, B. P., & Willemsen, M. C. (2015). Evaluating Recommender Systems with User Experiments. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook (pp. 309–352).* Springer US. https://doi.org/10.1007/978-1-4899-7637-6_9

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction, 22(4–5),* 441–504. https://doi.org/10.1007/s11257-011-9118-4

Knijnenburg, B. P., Willemsen, M. C., & Broeders, R. (2014). Smart Sustainability through System Satisfaction: Tailored Preference Elicitation for Energy-saving Recommenders. Human Computer Interaction.

Koestner, R., Otis, N., Powers, T. A., Pelletier, L., & Gagnon, H. (2008). Autonomous Motivation, Controlled Motivation, and Goal Progress. *Journal of Personality, 76(5),* 1201–1230. https://doi.org/10.1111/j.1467-6494.2008.00519.x

Kweku, D. W., Bismark, O., Maxwell, A., Desmond, K. A., Danso, K. B., Oti-Mensah, E. A., ... & Adormaa, B. B. (2018). Greenhouse effect: greenhouse gases and their impact on global warming. *Journal of Scientific research and reports*, *17*(6), 1-9. https://doi.org/10.9734/JSRR/2017/39630

Lee, J. W. C., & Tanusia, A. (2016, August). Energy conservation behavioural intention: Attitudes, subjective norm and self-efficacy. *In IOP conference series: Earth and environmental science (Vol. 40, No. 1, p. 012087).* IOP Publishing. (From: Open Access proceedings Journal of Physics: Conference series (iop.org))

Lesic, V., De Bruin, W. B., Davis, M. C., Krishnamurti, T., & Azevedo, I. M. (2018). Consumers' perceptions of energy use and energy savings: A literature review. *Environmental Research Letters*, *13(3)*, 033004. https://doi.org/10.1088/1748-9326/aaab92

Lindenberg, S., & Steg, L. (2007). Normative, Gain and Hedonic Goal Frames Guiding Environmental Behavior. *Journal of Social Issues, 63(1),* 117–137. https://doi.org/10.1111/j.1540-4560.2007.00499.x

Lindsey, R. (2022, April 19). Climate Change: Global Sea Level. NOAA Climate.gov. https://www.climate.gov/news-features/understanding-climate/climate-change-global-sea-level

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57(9),* 705–717. https://doi.org/10.1037/0003-066X.57.9.705

Locke, E. A., & Latham, G. P. (1990). A theory of goal setting & task performance. *Prentice-Hall, Inc.* https://doi.org/10.2307/258875

Lü, L., Medo, M., Yeung, C. H., Zhang, Y. C., Zhang, Z. K., & Zhou, T. (2012). Recommender systems. *Physics reports, 519(1),* 1-49. https://doi.org/10.1016/j.physrep.2012.02.006

Matsumoto, S., Mizobuchi, K., & Managi, S. (2021). Household energy consumption. *Environmental Economics and Policy Studies*, *24(1),* 1–5. https://doi.org/10.1007/s10018-021-00331-9

Milieu Centraal. (n.d.) . Gemiddeld energieverbruik in Nederland. Milieucentraal.nl. https://www.milieucentraal.nl/energie-besparen/inzicht-in-je-energierekening/gemiddeld-energieverbruik/#waar-hangt-je-verbruik-van-af

Milieu Centraal & Motivaction. (2023, March). Ook met prijsplafond willen mensen energie besparen, maar laten nog kansen liggen. Retrieved from https://www.milieucentraal.nl/persberichten/ook-met-prijsplafond-willen-mensen-energie-besparen/

Mousavi, M. E., Irish, J. L., Frey, A. E., Olivera, F., & Edge, B. L. (2011). Global warming and hurricanes: the potential impact of hurricane intensification and sea level rise on coastal flooding. *Climatic Change, 104*, 575-597. http://doi.org/10.1007/s10584-009-9790-0

Muthén, L. K., & Muthén, B. O. (2023). Mplus: Statistical Analysis with Latent Variables: User's Guide (Version 8.5). Los Angeles, CA: Authors.

NASA. (2022, September). Graphic: The relentless rise of carbon dioxide – Climate Change: Vital Signs of the Planet. https://climate.nasa.gov/climate_resources/24/graphic-the-relentless-rise-of-carbon-dioxide/

NASA. (2023, March 2). Climate Change Evidence: How Do We Know? Climate Change: Vital Signs of the Planet. https://climate.nasa.gov/evidence/#:~:text=In%201896%2C%20a%20seminal%20paper,temperature%20through%20the%20greenhouse%20effect.

Newton, P., & Meyer, D. (2013). Exploring the attitudes-action gap in household resource consumption: does "environmental lifestyle" segmentation align with consumer behaviour?. *Sustainability*, *5(3),* 1211-1233. https://doi.org/10.3390/su5031211

Ryan, T. A. (1970). Intentional behavior: An approach to human motivation.

Schäfer, H., & Willemsen, M. C. (2019). Rasch-based tailored goals for nutrition assistance systems. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 18–29. https://doi.org/10.1145/3301275.3302298

Schleussner, C. F., Rogelj, J., Schaeffer, M., Lissner, T., Licker, R., Fischer, E. M., ... & Hare, W. (2016). Science and policy characteristics of the Paris Agreement temperature goal. *Nature Climate Change, 6(9),* 827-835. https://doi.org/10.1038/NCLIMATE3096

Soper, D. S. (2020). A-priori sample size calculator for structural equation models [Software].

Stadelmann, M., & Schubert, R. (2018). How Do Different Designs of Energy Labels Influence Purchases of Household Appliances? A Field Study in Switzerland. Ecological Economics, 144, 112–123. https://doi.org/10.1016/j.ecolecon.2017.07.031

Starke, A. D. (2019). Supporting energy-efficient choices using Rasch-based recommender interfaces. [Doctoral dissertation, Eindhoven University of Technology]

Starke, A. D., Willemsen, M. C., & Snijders, C. C. P. (2017). Effective User Interface Designs to Increase Energy-efficient Behavior in a Rasch-based Energy Recommender System. Proceedings of the Eleventh ACM Conference on Recommender Systems, 65–73. https://doi.org/10.1145/3109859.3109902

Starke, A. D., Willemsen, M. C., & Snijders, C. C. P. (2020). Beyond "one-size-fits-all" platforms: Applying Campbell's paradigm to test personalized energy advice in the Netherlands. *Energy Research & Social Science, 59*, 101311. https://doi.org/10.1016/j.erss.2019.101311

Starke, A. D., Willemsen, M. C., & Snijders, C. C. P. (2021). Promoting Energy-Efficient Behavior by Depicting Social Norms in a Recommender Interface. *ACM Transactions on Interactive Intelligent Systems, 11(3–4),* 1–32. https://doi.org/10.1145/3460005

Tiefenbeck, V., Tasic, V., Schöb, S., & Staake, T. (2016). Long-lasting effects or short-term spark? On the persistence of behaviour change induced by real-time feedback on resource consumption.

Umit, R., Poortinga, W., Jokinen, P., & Pohjolainen, P. (2019). The role of income in energy efficiency and curtailment behaviours: Findings from 22 European countries. *Energy Research & Social Science, 53*, 206–214. https://doi.org/10.1016/j.erss.2019.02.025

Ungemach, C., Camilleri, A. R., Johnson, E. J., Larrick, R. P., & Weber, E. U. (2018). Translated Attributes as Choice Architecture: Aligning Objectives and Choices Through Decision Signposts. *Management Science*, *64(5)*, 2445–2459. https://doi.org/10.1287/mnsc.2016.2703

United Nations (2022). Net Zero Coalition | United Nations. United Nations; United Nations. https://www.un.org/en/climatechange/net-zero-coalition

Willemsen, M. C., Graus, M. P., & Knijnenburg, B. P. (2016). Understanding the role of latent feature diversification on choice difficulty and satisfaction. *User Modeling and User-Adapted Interaction, 26(4),* 347–389. https://doi.org/10.1007/s11257-016-9178-6

Wood, R. E., Mento, A. J., & Locke, E. A. (1987). Task complexity as a moderator of goal effects: A meta-analysis. *Journal of applied psychology, 72(3),* 416. https://doi.org/10.1037/0021-9010.72.3.416

World Health Organization. (n.d.). *Drought*. Who.int; World Health Organization: WHO. Retrieved March 9, 2023, from https://www.who.int/health-topics/drought

Zhang, Y., Bai, X., Mills, F. P., & Pezzey, J. C. (2021). Examining the attitude-behavior gap in residential energy use: empirical evidence from a large-scale survey in Beijing, China. *Journal of Cleaner Production*, *295*, 126510. https://doi.org/10.1016/j.jclepro.2021.126510

Appendix A: Energy Saving Measures.

ID	Name	Difficulty	Energy savings in kWh	Gas Or Electricity
1	Place brushes/valves on mailboxes and/or keyholes.	-0.7	25	Gas
2	Ensure good air circulation in your refrigerator.	-0.9	2	Electricity
3	Throw away large food scraps, but don't pre-wash.	-0.3	2	Gas
4	Clean lamps and their sockets regularly.	1.81	2	Electricity
5	Clean the coils at the back of your refrigerator regularly.	2.97	15	Electricity
6	Regularly replace/clean the filters or bag of your vacuum cleaner.	0	5	Electricity
7	Turn off the oven 10 minutes earlier.	2.63	10	Electricity
8	Set boiler temperature to 60-65°C.	0.15	900	Gas
9	Save your laundry.	-3.6	30	Electricity
10	Turn off the air conditioning in unused rooms.	-0.9	100	Electricity
11	Turn off the ventilation in your bathroom for a maximum of 20 minutes after showering.	-1.2	12	Electricity
12	Insulate your water heater so it feels cold to the touch.	0.8	130	Gas
13	Insulate your attic floor, including attic door or hatch.	1.25	6400	Gas
14	Replace light bulbs with energy saving bulbs or better.	-2	140	Electricity
15	Install windows with double glazing or better.	-1.6	2500	Gas
16	Place light switches at both ends of a hallway.	-1.3	17	Electricity
17	Use an energy-efficient monitor for your desktop computer.	0.37	20	Electricity

18	Replace incandescent and energy-saving lamps with LED lamps.	-1.7	30	Electricity
19	Buy a laptop or mini PC instead of a desktop computer.	-2.4	40	Electricity
20	Do not boil more water than you need.	-1.4	30	Electricity
21	Apply weatherstripping to doors.	0	120	Gas
22	Turn off your PC when you are not using it.	-0.6	100	Electricity
23	Buy an energy-efficient washing machine.	-0.3	200	Electricity
24	Install a shower water consumption meter.	2.89	200	Gas
25	Clean up leaves with a rake instead of a leaf blower.	-1.8	36	Electricity
26	Install solar garden lights.	1.72	50	Electricity
27	Thaw food in the refrigerator or sink, rather than in the microwave.	-1.1	10	Electricity
28	Prune plants with a non-electric pruning or hedge trimmer	-0.1	5	Electricity
29	Buy a hand lawnmower instead of a motorized lawnmower.	2.1	50	Electricity
30	Set your TV to be energy efficient.	1.1	50	Electricity
31	Insulate your cavity walls.	0	8000	Gas
32	Let food cool before putting it in the fridge.	-2.3	2	Electricity
33	Place rugs on your floor.	0.14	60	Gas
34	Wash at a low(er) temperature.	-1.4	105	Electricity
35	Use a clothesline or drying rack instead of the dryer.	-2.2	290	Electricity
36	Apply draft strips to the windows	-0.5	460	Gas
37	Seal cracks in exterior walls.	-0.1	250	Gas
38	Do not use a hairdryer to dry your hair.	-1	30	Electricity
39	Place radiator foil (behind radiators).	1.96	900	Gas
40	Set your thermostat to a lower temperature if you are going away for a few days.	-2.3	115	Gas

41	Cook on an induction hob instead of a gas stove.	1.6	185	Gas
42	Install water-saving aerators on all your taps.	0.58	500	Gas
43	Shower short(er).	0.64	185	Gas
44	Turn off your water heater if you are going away for a few days.	1.01	20	Gas
45	Turn off the tap while you soap yourself.	1.16	400	Gas
46	Use a broom instead of a vacuum cleaner.	0.28	25	Electricity
47	Repair leaky taps.	-2.1	70	Gas
48	Only run the dishwasher when it is full.	-2.2	40	Electricity
49	Use a skillet instead of the oven.	-1.1	250	Electricity
50	Cook with the correct size pan on the corresponding burner.	-2.7	5	Gas
51	Use a lid when cooking.	-2.2	5	Gas
52	Decide what you are going to take out of the fridge/freezer before you open it.	-1.7	40	Electricity
53	Set your refrigerator to 4°C.	-1.2	50	Electricity
54	Set your freezer to -18°C.	-0.4	50	Electricity
55	Install a programmable thermostat.	0.48	750	Gas
56	Install a remote thermostat.	1.77	1000	Gas
57	Insulate ceilings.	0.79	7400	Gas
58	Insulate heating pipes.	0.3	100	Gas
59	Insulate roofs.	0.08	8000	Gas
60	Do not use an electric heater to heat large spaces.	-0.8	800	Electricity
61	Seal holes in insulation material with polyurethane foam.	0.17	1000	Gas
62	Lower the thermostat by 1°C.	-0.5	1100	Gas
63	Set the thermostat to 14°C before going to sleep.	-0.4	1250	Gas
64	Put timers on patio heaters.	2.99	50	Electricity
65	Turn off the air conditioner when no one is home.	-0.5	50	Electricity

66	Use an extra blanket instead of the heater.	-1.7	1000	Gas
67	Use an electric blanket instead of the heater.	1.69	740	Gas
68	Buy an energy-efficient freezer.	0.38	190	Electricity
69	Install a heat recovery system in the shower.	3.59	1100	Gas
70	Install a water-saving shower head.	0.22	400	Gas
71	Install motion sensors on your lamps.	0.89	25	Gas
72	Install energy-efficient fixtures.	0.38	50	Electricity
73	Use daylight and leave lights off as much as possible during the day.	-1.4	150	Electricity
74	Dry only full loads in the dryer.	-1.2	50	Electricity
75	Install an energy consumption manager.	2.72	750	Electricity
76	Buy an energy-efficient dishwasher.	0	70	Electricity
77	Enable power management of your desktop PC or laptop.	-0.1	40	Electricity
78	Install a heat pump to heat the house.	2.66	3100	Gas
79	Do not use an air conditioner.	-2.1	400	Electricity
80	Renew your old water heater.	1.24	800	Gas
82	Take a shower instead of taking a bath.	-1.9	400	Gas
83	Do not use an electric blanket, but another blanket.	-1.8	60	Electricity
84	Turn off the coffee maker completely.	-2	80	Electricity
85	Air chambers 20 minutes daily.	-0.7	250	Gas
86	Wash by hand instead of in the dishwasher.	-0.2	160	Electricity
87	Stir fry	-0.2	20	Gas
88	Make coffee without a hot plate.	0.14	25	Electricity
89	Air clothes instead of washing them.	0.11	30	Electricity
90	Turn off the dishwasher immediately after use.	-1.3	70	Electricity

91	Turn off the washing machine completely.	0.46	95	Electricity
92	Turn off the PC with switch box.	0	175	Electricity
93	Buy an energy-efficient fridge/freezer combination.	-0.1	160	Electricity
94	Keep extractor hood filter(s) clean.	0.89	2	Electricity
95	Green power	0.26	0	Electricity
96	Descale your coffee maker and/or kettle.	0.39	10	Electricity
97	Free the fridge/freezer from ice.	0.03	50	Electricity
98	Tumble dry shirts briefly instead of ironing them.	0.63	25	Electricity
99	Carry out maintenance on the central heating boiler or geyser.	0	120	Gas
100	Day-night rate	0.36	0	Electricity
101	Lower the boiler temperature.	0.97	40	Gas
102	Replace dimmers.	-0.1	60	Electricity
103	Descale the washing machine.	1.53	10	Electricity
104	Buy an energy-efficient dryer.	0.81	210	Electricity
105	Install door closers.	1.48	220	Gas
106	Turn off the refrigerator when on vacation.	0.98	20	Electricity
107	Install solar panels.	1.65	2000	Electricity
108	Wash with a hot-fill washing machine.	3.11	40	Electricity
109	Install a solar water heater.	3.01	1850	Gas
110	Place a mini windmill.	5.66	1400	Electricity
111	Iron several garments at a time.	-0.7	2	Electricity
112	Use rechargeable batteries.	-0.8	0	Electricity
113	Install a timer on your water heater.	2.21	500	Gas
114	Bleed your radiators regularly.	1.04	300	Gas
115	Use your tablet instead of your laptop/desktop.	0.86	80	Electricity
116	Get rid of your second refrigerator.	-0.3	240	Electricity
117	Use the eco program of your dishwasher.	-0.6	85	Electricity
118	Install floor insulation.	0.35	2450	Gas
119	Install thermostatic mixing valves.	-0.3	70	Gas

120	Install exterior wall insulation.	1.52	1700	Gas
121	Make sure that the rubber seals of your refrigerator are kept clean and airtight.	0.53	25	Electricity
122	Use an energy-efficient TV.	-0.1	120	Electricity
123	Place your refrigerator in the right place.	0.1	25	Electricity
124	Open and close curtains and the like at the right times.	-1.3	1000	Gas
125	Place deciduous trees around your home.	0.5	1000	Gas
126	Unplug devices and chargers when not in use.	-0.3	25	Electricity
127	Turn off screens when not in use.	-1.5	45	Electricity
128	Turn off your computer screen while downloading.	0.99	6	Electricity
129	Cook in a pressure cooker.	3.32	30	Gas
130	Don't use a screen saver.	-1.7	20	Electricity
131	Do not leave your extractor hood on for an unnecessarily long time.	-2.7	18	Electricity
132	Install underfloor heating (if your house is well insulated).	0.7	640	Gas
133	Watch one hour less TV per day.	0.57	75	Electricity
134	Check the pressure in your boiler.	0.63	0	Gas
135	Install a smaller cistern on your toilet(s).	0.29	0	

Appendix B: Saving Aid Workings.

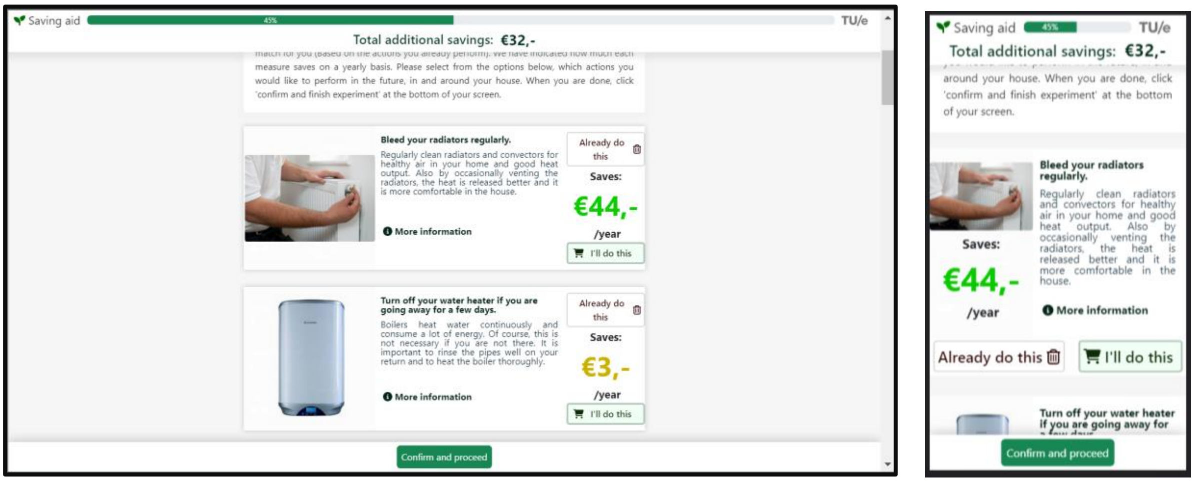
App contents and logic. The Saving Aid app was made with React as front end, and a PHP API as a back end. It used SQL databases for storing measures, user data, questions and several parameters including kWh and gas prices. Requests from React to the API to get, update or post data, were made with the React AJAX library, Axios.

React. React is a single-page, component-based, front-end framework, especially useful for websites that show items with the same styling and behavior, but different contents, e.g., news websites with different articles. It allows you to define logic and styling for a component once, and then re-render this component with varying content but identical logic and styling (e.g., the measurements, of which there were 135). This content can be stored in table format for which every row is a new object; either as a JSON, excel or other format (in our case SQL). Each component can then be populated with the data of a particular row of this database. In our case, simultaneously styling 135 measures was done by developing a single 'new action' component with react and re-populating this with data from the measures we wanted to display.

Components. The app consisted of one main component, rendering each page in a sequential fashion, dependent on the value of the 'currentPage' variable. This ensured that participants could not go back to a previous page. The sequence of pages looked as follows (figure below), with users initially seeing the pages from "*pick language*" until the "*thank you*" page, and the follow-up pages 3 to 4 weeks after. (With additional consent and thank you components). Participants who had a goal condition viewed goal setting and instruction pages, whereas other participants skipped over these pages. Registration and login functionality were eventually not fully implemented. With URL parameters, participant IDs could be captured and in the current system, this made registration unnecessary.

In the above development menu, different pages, conditions, participant types, and languages could be tested during development.

Front end. The front end was based on screenshots from earlier research, paying closer attention to the responsiveness of the website such that it could be used on mobile. Below you can see the difference in layout between the desktop interface on the left and the mobile interface on the right.



State variables in Redux. Usually, a React component keeps state variables, e.g. in our case, we might want to keep track of when a user has selected a certain measure, or unselected it. Then, we created a state variable 'selected'. When this state is changed, the component re-renders, allowing you to apply conditional styles. However, when one component (for example the navigation bar) needs to know the state of another, or multiple components, states need to be communicated between them. In our case, showing the number of selected items in the navigation bar would require each component to have access to this number, and communicate state changes back up. When multiple such information streams are needed, this communication can become rather complex. Redux offers a solution to this by maintaining a centralized 'store', where state variables can be stored, and accessed easily by any component.

For keeping track of user actions on various pages, a single 'user data' object was maintained in the store, that kept track of what ability a user has, how many items they have selected, how much kWh that amounted to in total, etc. Each time a user went to a new page, the entire object was communicated to the back end via Axios.

Language. This state object also stored the language picked by a user (Either English or Dutch). All buttons, instructions, measures, and questions, were made with an English and a Dutch version that were displayed dependent on this variable.

Review and follow up. For the review and follow-up, we sent participants a link that included their random id (30 character randomly generated ID). They were sent to a redirect PHP page where we could read this id and retrieve information from a separate database on their chosen measures, language, and signpost condition. After this, participants saw their chosen measures with the correct signpost and in the correct language on the React website. This follow-up database had no sensitive data.

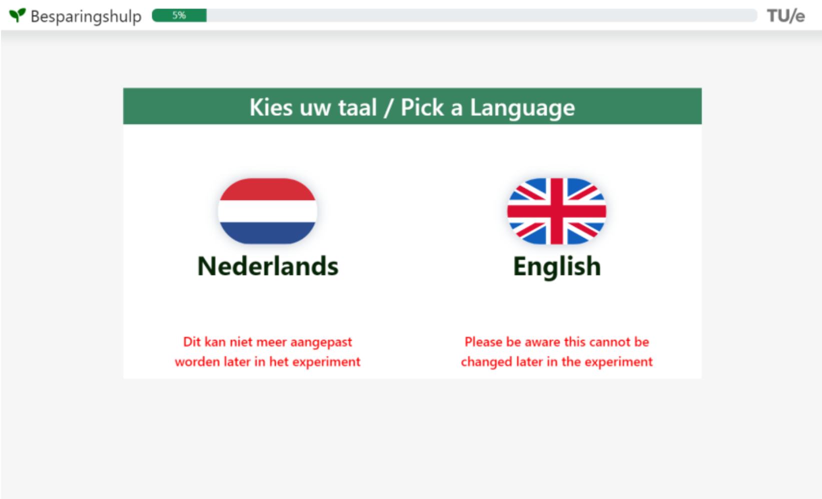
Security. To ensure data protection, CORS access was restricted such that only our domain could communicate with the API. Furthermore, to prevent SQL injection, all SQL queries used bound parameters that were inserted into existing SQL queries, thus preventing manual, malicious queries from being generated by a user. Additionally, user data could only be posted or adapted through the API, but not retrieved as this was not necessary for our application to work. User data was only obtained through the PHP My-Admin panel from the hosting provider, which requires experimenter login. Furthermore, a GDPR complaint, European Union-based hosting provider, was used for this study.

Appendix C: Saving-Aid Screenshots In Order Of Experiment

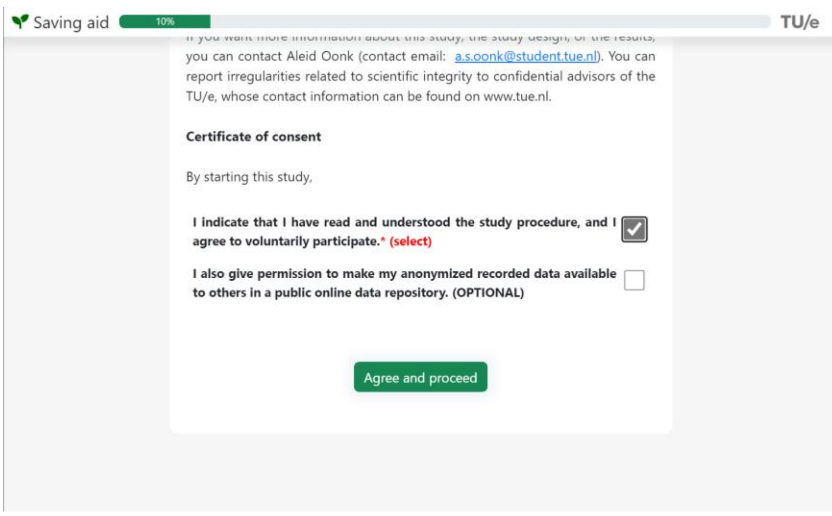
Desktop on the left, mobile on the right.

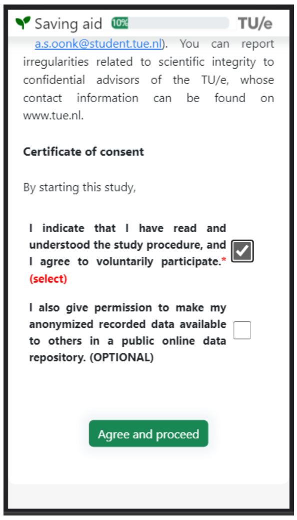
Initial study

Screen 1: Language selection

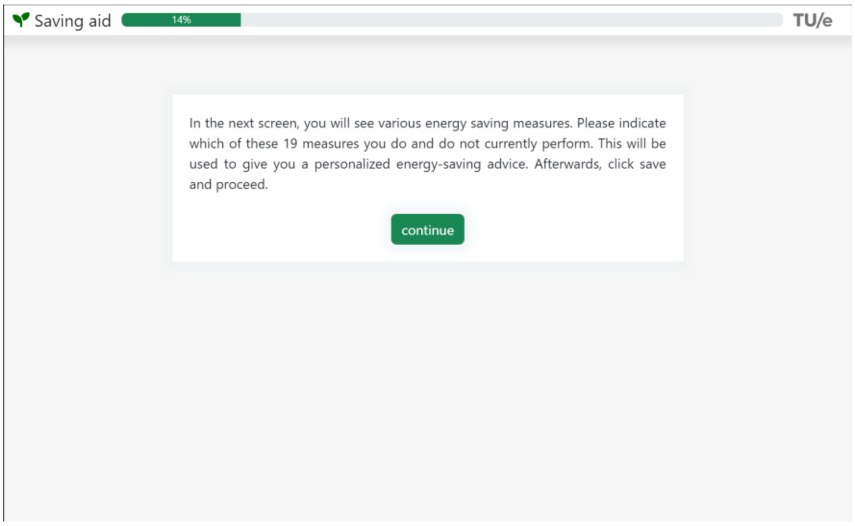


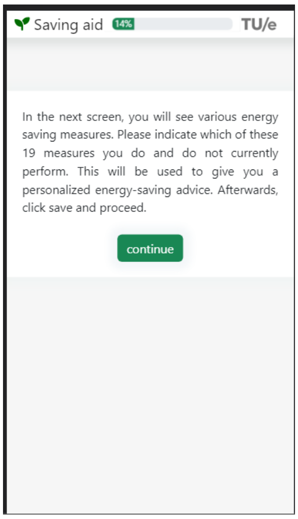
Screen 2: Consent



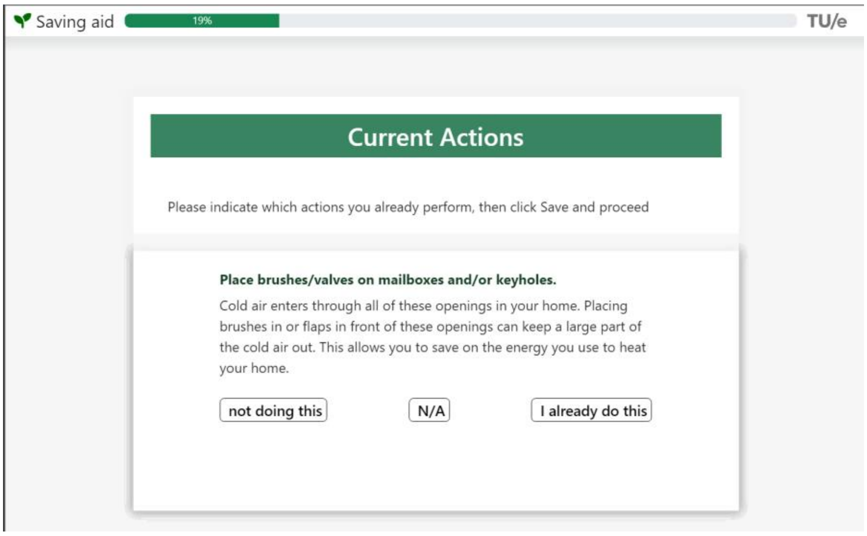


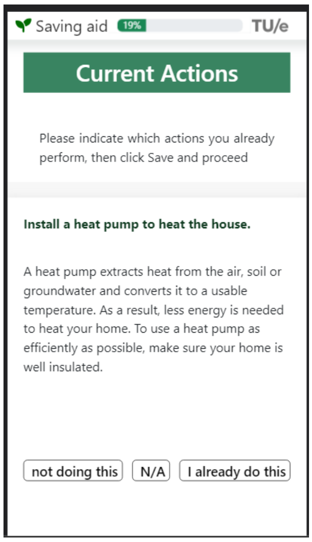
Screen 3: current actions instructions



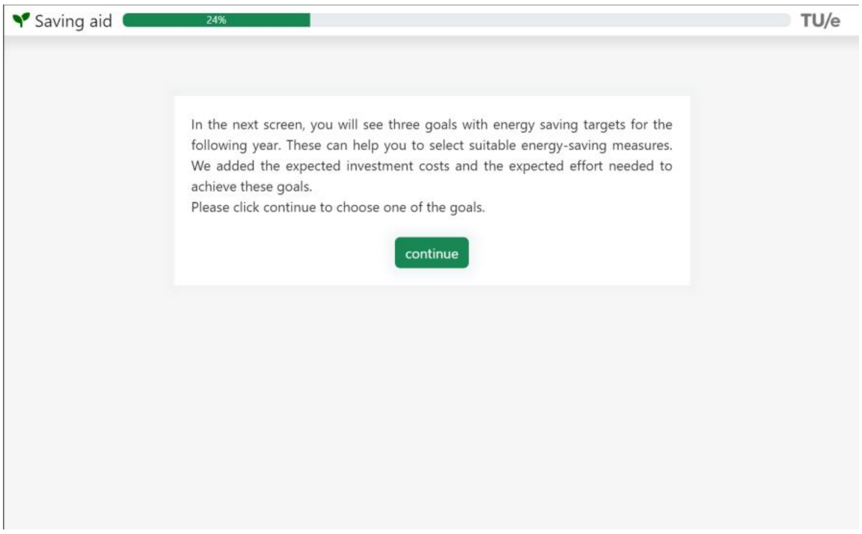


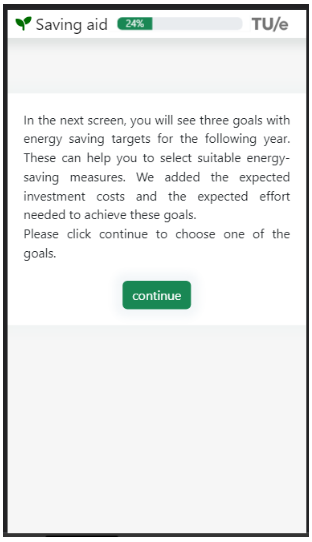
Screen 4: current actions





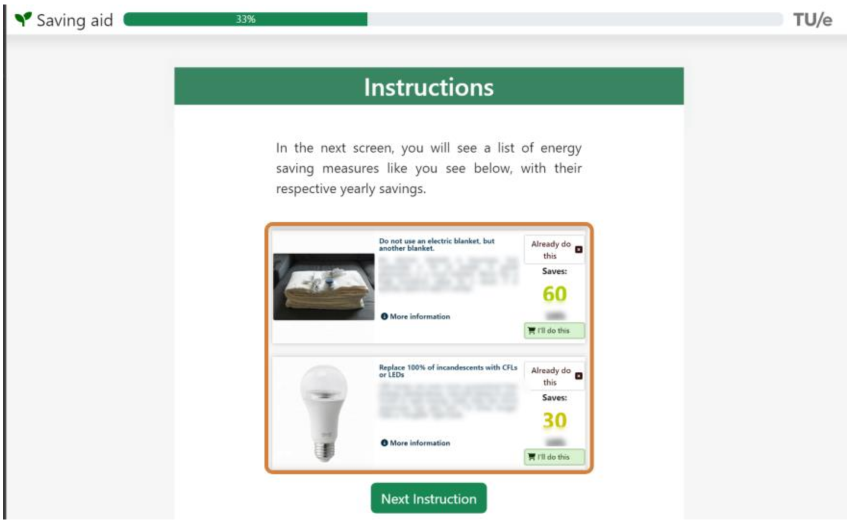
Screen 5: goal-setting instructions (only for goal condition)

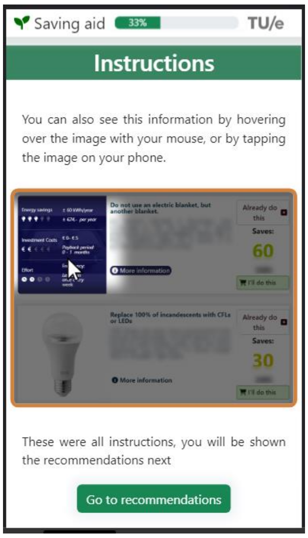




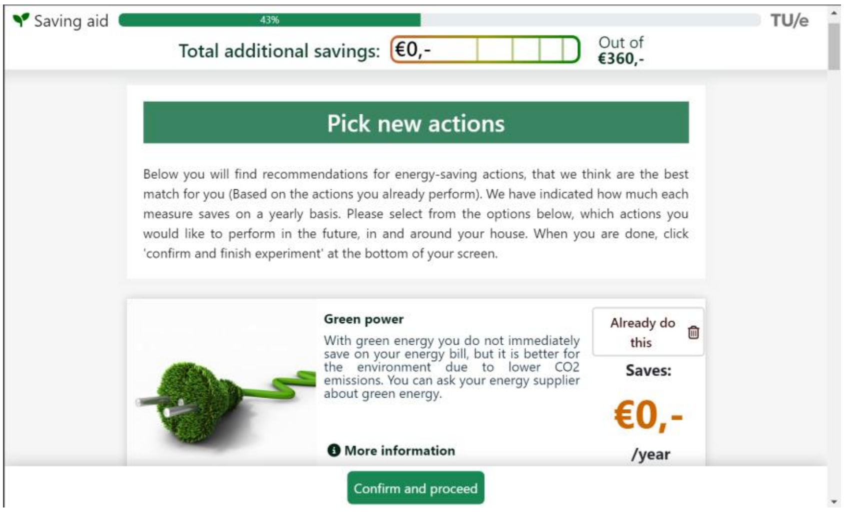
Screen 6: goal setting (only for goal condition)

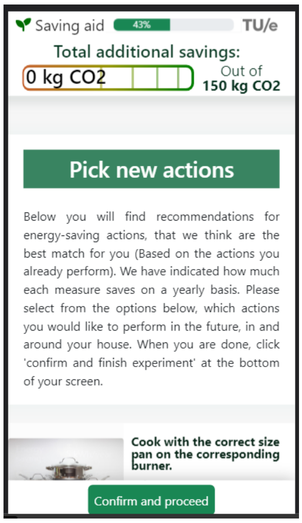
Screen 7: new actions instructions (5 total, as seen in Figure 6)

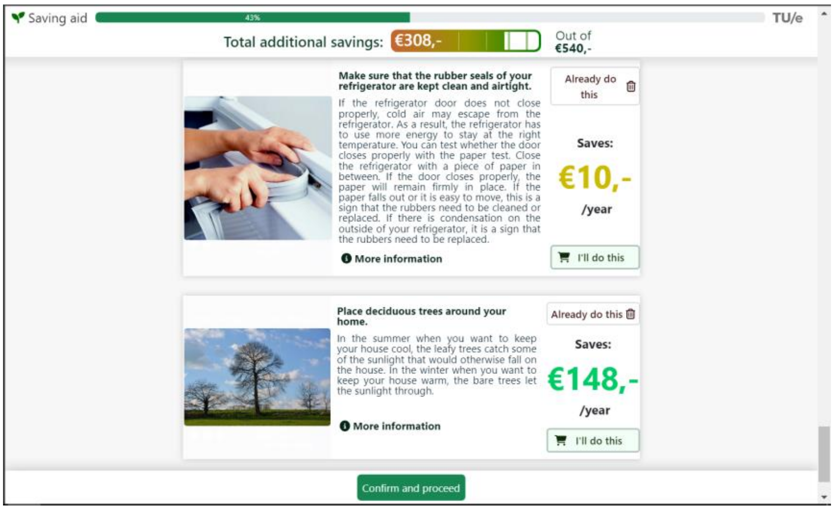


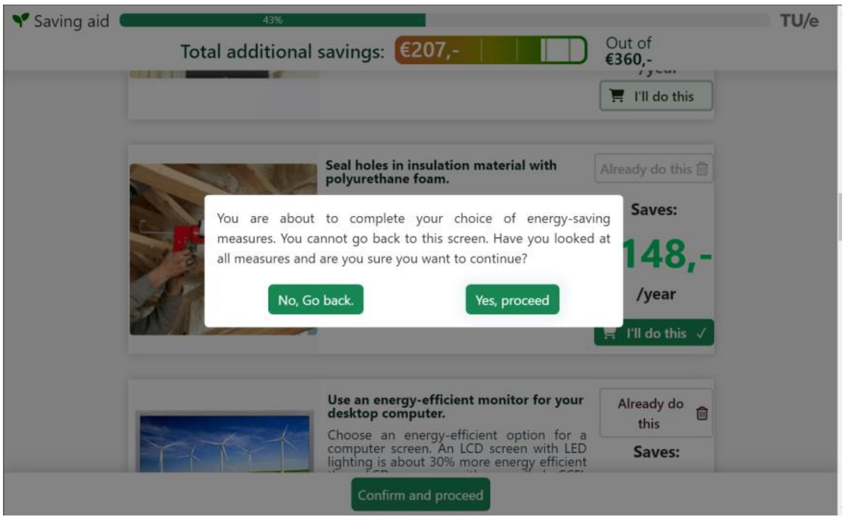


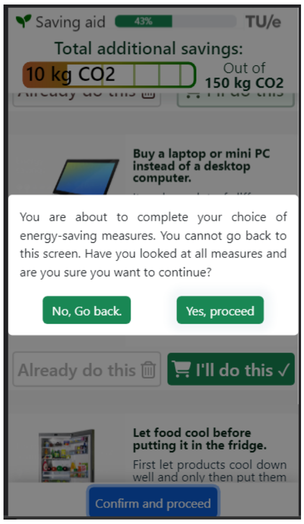
Screen 8: new actions



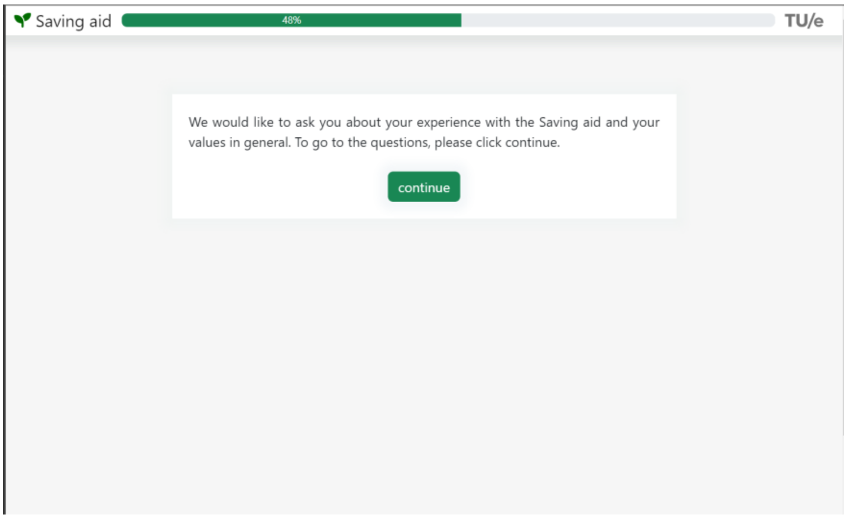


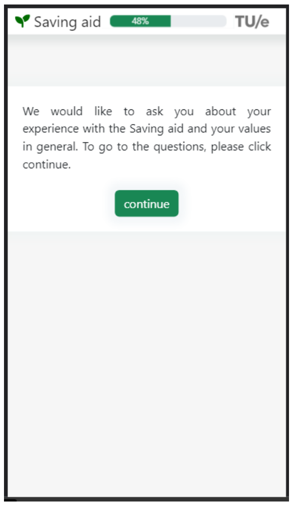




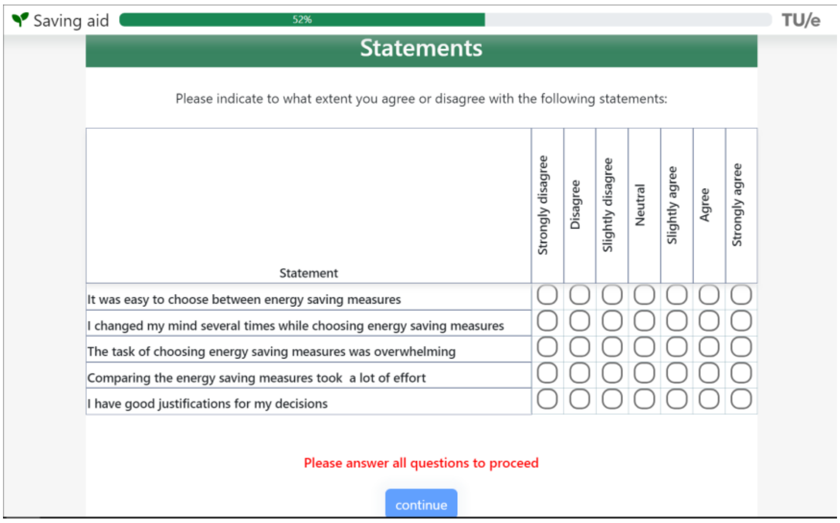


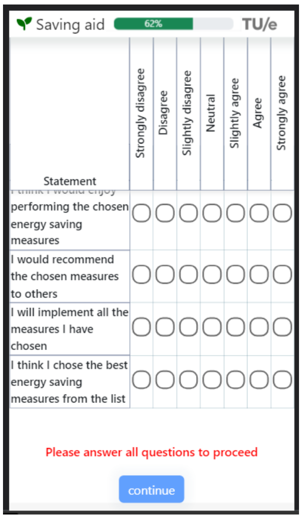
Screen 9: Questionnaire instructions



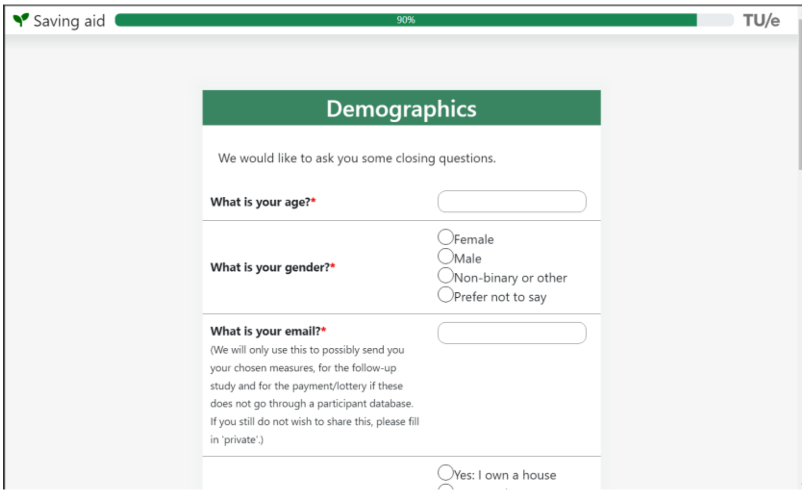


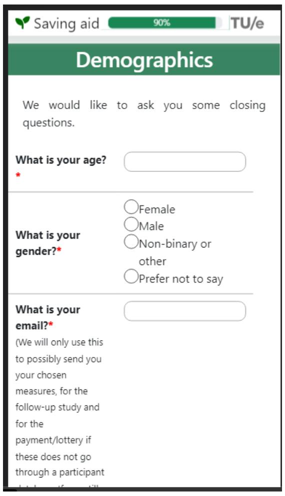
Screen 10-17: Questionnaires (choice difficulty, perceived feasibility, choice satisfaction, system satisfaction, goal support, environmental self-efficacy, nep scale, and IMS scale)



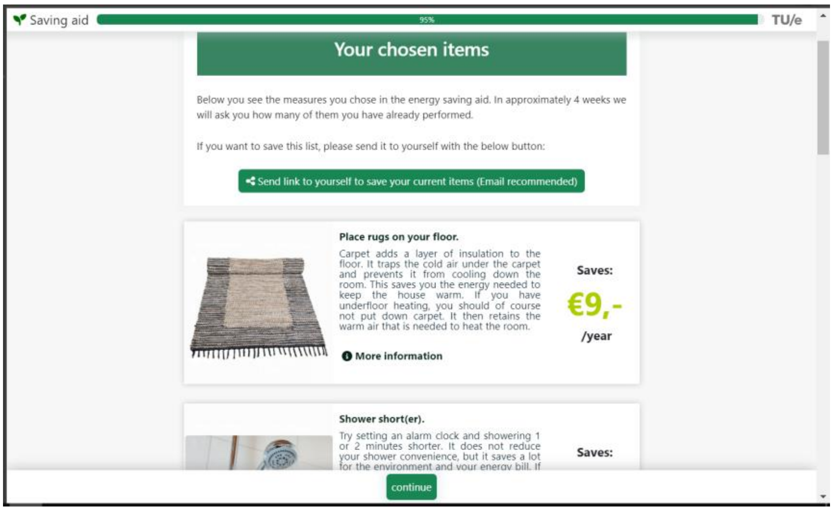


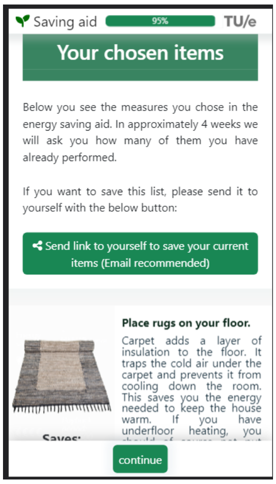
Screen 18: Demographics



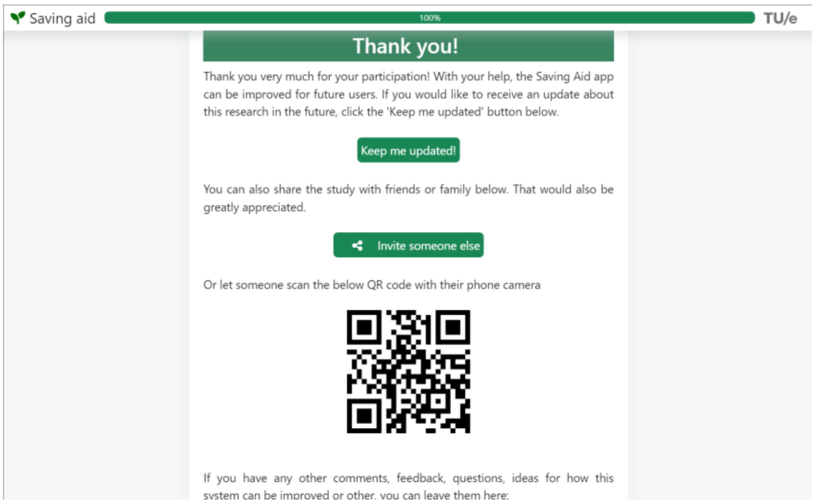


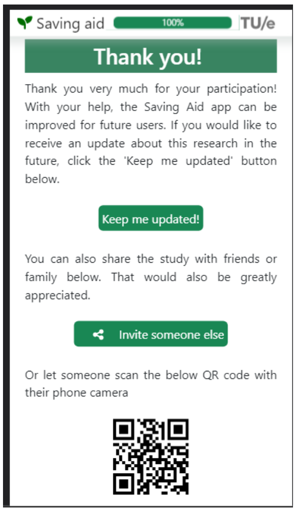
Screen 19: Review items



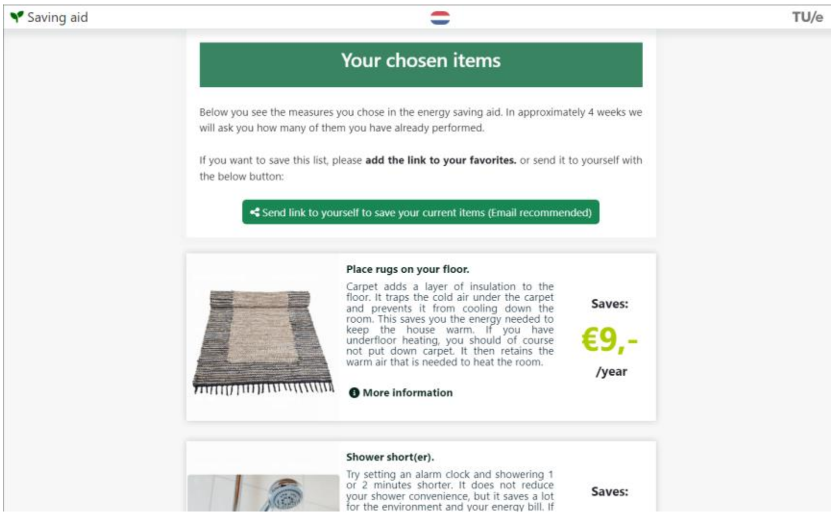


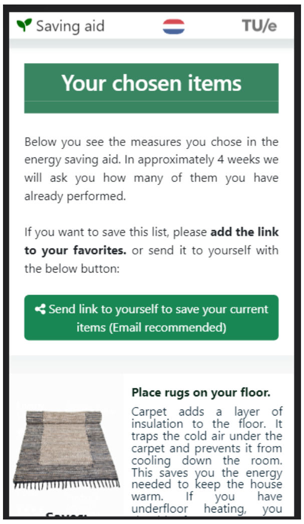
Screen 20: Thank you



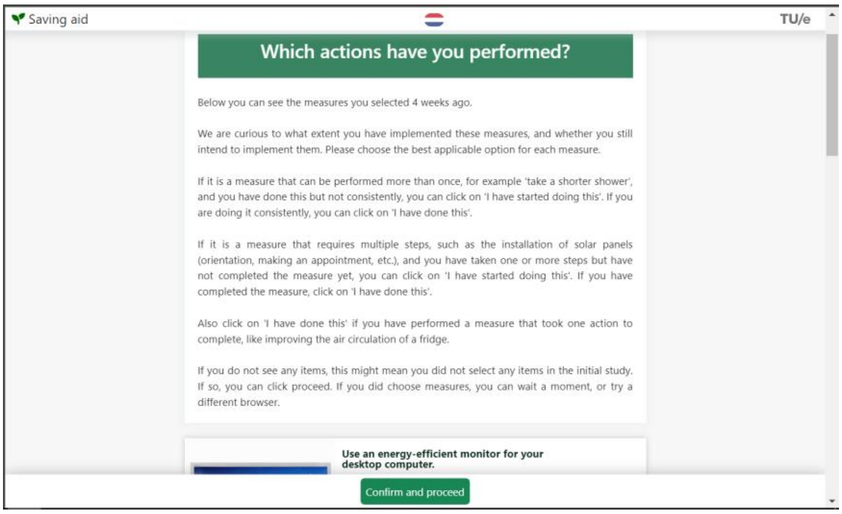


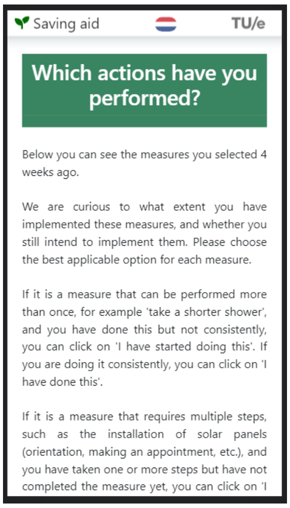
Review



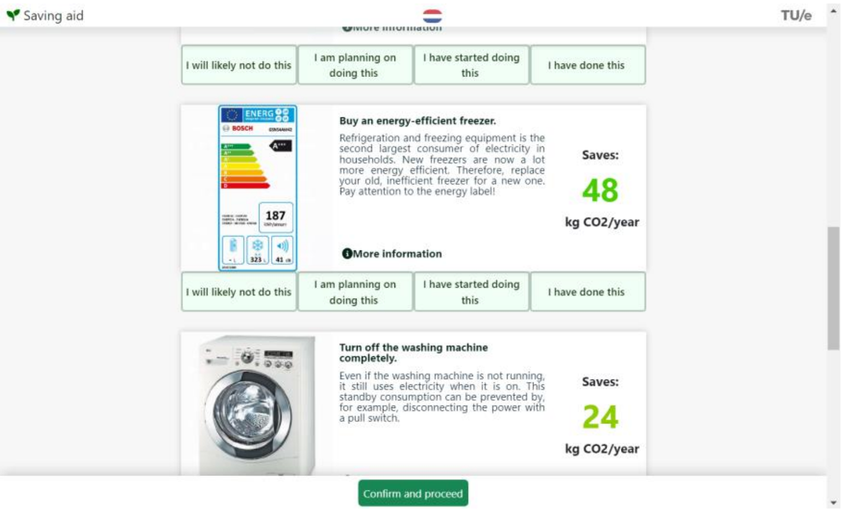


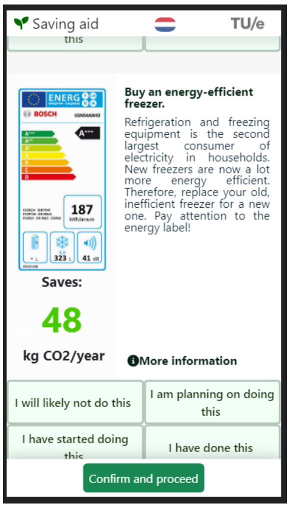
Follow-up study





(Items collapsed and disappeared on clicking one of the buttons)





Appendix D: Questionnaires

Choice difficulty (Willemsen et al., 2016) - **α** 0.73 **AVE:** 0.555

id	*name*	*question*	*questionNL*
1	CDIF1	It was easy to choose between energy saving measures	Het was gemakkelijk om te kiezen tussen energiebesparende maatregelen
2	CDIF2	I changed my mind several times while choosing energy saving measures	Ik ben verschillende keren van gedachten veranderd tijdens het kiezen van energiebesparende maatregelen
3	CDIF3	The task of choosing energy saving measures was overwhelming	De taak om energiebesparende maatregelen te kiezen was overweldigend
4	CDIF4	Comparing the energy saving measures took a lot of effort	Het vergelijken van energiebesparende maatregelen kostte veel moeite
5	CDIF5	I have good justifications for my decisions	Ik heb een goede onderbouwing voor mijn beslissingen

- ***7 point Likert scale disagree/agree***

Perceived feasibility (Starke et al., 2017) **α**: 0.83 (study 2)

id	*name*	*question*	*questionNL*
1	PF1	I think it would take me little effort to perform the chosen measures.	Het zou me denk ik weinig moeite kosten om de gekozen maatregelen uit te voeren.
2	PF2	I do not have the possibility to perform the chosen measures.	Ik heb niet de mogelijkheid om de gekozen maatregelen uit te voeren.
4	PF3	I think the chosen measures are easy to apply in my home environment.	De gekozen maatregelen zijn denk ik makkelijk toe te passen in mijn thuisomgeving.

- ***7 point Likert scale disagree/agree***

Choice satisfaction (From Willemsen et al., 2016 (**α:** 0.93 **AVE**: 0.847) and Knijnenburg et al., 2014 (Study 4: **AVE:** 0.538))

id	name	*question*	*questionNL*	*Source(s)*
1	Chsat1	I am satisfied with the measures I chose	Ik ben tevreden met de maatregelen die ik heb gekozen	Willemsen16, ~Knijnenburg14
2	Chsat2	I think I would enjoy performing the chosen energy saving measures	Ik denk dat ik het leuk zou vinden om de gekozen maatregelen uit te voeren	Willemsen16, ~Knijnenburg14
3	Chsat3	I would recommend the chosen measures to others	Ik zou de gekozen maatregelen aanbevelen aan anderen	Willemsen16
4	Chsat4	I will implement all the measures I have chosen	Ik ga alle maatregelen uitvoeren die ik heb gekozen.	~KN14 ('How many measures will you implement?)
5	Chsat5	I think I chose the best energy saving measures from the list	Ik denk dat ik de beste energie-besparings maatregelen uit de lijst heb gekozen.	Knijnenburg14, Willemsen16

7 point Likert scale disagree/agree

System satisfaction scale, with questions from System satisfaction (Knijnenburg et al., 2014), system effectiveness (Knijnenburg et al., 2012) and perceived support (Starke et al., 2017) scales, See next page for original questionnaires.

id	name	*question*	*questionNL*	source
1	SYSSAT1	I make better choices using the Saving aid.	Met de Besparingshulp maak ik betere keuzes.	Starke 2017, Knijnenburg 2012
2	SYSSAT2	The Saving aid is helpful to find appropriate measures.	De Besparingshulp helpt bij het vinden van passende maatregelen.	Starke 2017

3	SYSSAT3	I would recommend the Saving aid to others	Ik zou de Besparingshulp aan anderen aanbevelen	Knijnenburg 2012 & 2014
4	SYSSAT4	I would use the Saving aid more often if possible	Indien mogelijk zou ik de Besparingshulp vaker gebruiken	Knijnenburg 2014 and Starke 2017
5	SYSSAT5	The Saving aid was useless	De Besparingshulp was waardeloos	all 3 papers

7 point Likert scale disagree/agree

Goal support: (leaving out the world 'goal' and directly focusing on energy saving behavior that will lead to that goal).

id	name	*question*	*questionNL*	*source*
1	GSUP1	The Saving aid makes saving energy easier	De Besparingshulp maakt het besparen van energie makkelijker	Own
2	GSUP2	The Saving aid motivates me to save more energy.	De Besparingshulp motiveert mij meer energie te besparen.	Own
3	GSUP3	I think I will save more energy in the coming year thanks to the Saving aid	Ik denk dat ik het komende jaar meer energie zal besparen dankzij de Besparingshulp	Own
4	GSUP4	The Saving aid gives me more insight into the energy consumption of devices and systems in my home	De Besparingshulp geeft mij meer inzicht in het energieverbruik van apparaten en systemen in mijn huis	Own
5	GSUP5	The Saving aid made me more energy-conscious	De Besparingshulp heeft me energiebewuster gemaakt	System satisfaction (Knijnenburg 2014)
6	GSUP6	The Saving aid makes me more aware of my options for saving energy	De Besparingshulp maakt me meer bewust van mijn mogelijkheden om energie te besparen	System effectiveness, (Knijnenburg 2012)

7 point Likert scale disagree/agree

Original questionnaires goal support / system satisfaction:

- **System satisfaction** (Knijnenburg et al., 2014)
- **Perceived support** (Starke et al., 2017) **α:** 0.81(study1) - 0.92(study2)
- **System effectiveness** (Knijnenburg et al., 2012)

Self efficacy scale regarding energy saving by Lee and Tanusia (2016), who adapted this from the 'general self-efficacy scale' by Chen et al. (2001)

id	*name*	*question*	**questionNL**
1	SEF1	I will be able to achieve most of the goals that I have set for myself concerning energy conservation.	De meeste besparingsdoelen die ik mezelf heb gesteld, kan ik bereiken
2	SEF2	When facing difficult decisions on energy conservation, I am certain that I will accomplish them.	Moeilijke beslissingen over energiebesparing, zal ik zeker kunnen nemen.
3	SEF3	In general, I think I can obtain energy conservation outcomes that are important to me.	Ik denk dat ik de besparingsresultaten kan behalen die voor mij belangrijk zijn.
4	SEF4	I believe I can succeed at most any energy conservation endeavour to which I set my mind.	Ik geloof dat ik in bijna elk besparingsstreven kan slagen, waar ik mijn zinnen op zet.
5	SEF5	I am confident that I can perform effectively on many different tasks relating to energy conservation.	Ik ben ervan overtuigd dat ik veel verschillende besparingstaken effectief kan uitvoeren.
6	SEF6	Compared to other people, I can do most energy conservation task very well	Vergeleken met andere mensen kan ik de meeste besparings-maatregelen heel goed uitvoeren

7 point Likert scale disagree/agree

Environmental attitudes: Revised NEP scale by Dunlap et al., 2000 **α:** 0.83.

id	*name*	*question*	**questionNL**
1	NEP1	We are approaching the limit of the number of people the earth can support	We bereiken bijna de grenzen van de hoeveelheid mensen die op aarde onderhouden kunnen worden.
2	NEP2	Humans have the right to modify the natural environment to suit their needs	Mensen hebben het recht om de natuurlijke omgeving te veranderen zodat hun eigen behoeften worden vervuld.
3	NEP3	When humans interfere with nature it often produces disastrous consequences	Als mensen ingrijpen in de natuur, heeft dat vaak rampzalige gevolgen.
4	NEP4	Human ingenuity will ensure that we do NOT make the earth unliveable	De vindingrijkheid van de mens zal ervoor zorgen dat we de aarde niet onleefbaar zullen maken.
5	NEP5	Humans are severely abusing the environment	De mens is het milieu ernstig aan het misbruiken.
6	NEP6	The earth has plenty of natural resources if we just learn how to develop them	De aarde heeft meer dan genoeg natuurlijke voorraden, we moeten alleen nog leren hoe we ze kunnen ontwikkelen.
7	NEP7	Plants and animals have as much right as humans to exist	Planten en dieren hebben evenveel recht om te bestaan als mensen.
8	NEP8	The balance of nature is strong enough to cope with the impacts of modern industrial nations	De balans van de natuur is sterk genoeg om met de gevolgen van de moderne industrielanden om te gaan.
9	NEP9	Despite our special abilities humans are still subject to the laws of nature	Ondanks onze bijzondere vaardigheden als mens, zijn we nog steeds onderworpen aan de wetten van de natuur.
10	NEP10	The so-called "ecological crisis" facing humankind has been greatly exaggerated	De zogenaamde 'ecologische crisis' die de mensheid boven het hoofd hangt, is sterk overdreven.

11	NEP11	The earth is like a spaceship with very limited room and resources	De aarde is net een ruimteschip met beperkte ruimte en beperkte middelen.
12	NEP12	Humans were meant to rule over the rest of nature	De mensheid is geschapen om over de rest van de natuur te heersen.
13	NEP13	The balance of nature is very delicate and easily upset	Het evenwicht van de natuur is erg gevoelig en gemakkelijk te verstoren.
14	NEP14	Humans will eventually learn enough about how nature works to be able to control it	Mensen zullen uiteindelijk genoeg leren over de werking van de natuur, dat zij in staat zullen zijn haar te beheersen.
15	NEP15	If things continue on their present course, we will soon experience a major ecological catastrophe	Als alles doorgaat op de manier waarop het nu gaat, zullen we snel een enorme ecologische catastrofe tegemoet gaan.

"eight odd-numbered items and disagreement with the seven even-numbered items indicate pro-NEP responses. ***cSA = Strongly Agree, MA = Mildly Agree, U = Unsure, MD = Mildly Disagree, and SD = Strongly Disagree"****(Dunlap et al., 2000)*

Monetary values: IMS: **Importance of Money scale**; by Franzen and Mader (2022) **α**: 0.82 (Paper only cited one time, but very recent study)

id	***name***	***question***	***questionNL***
1	IMS1	Material wealth is important for me.	Materiële rijkdom is belangrijk voor mij.
2	IMS2	Money is important for me.	Geld is belangrijk voor mij.
3	IMS3	Money makes me happy.	Geld maakt me gelukkig.
4	IMS4	Financial security is important for my well-being.	Financiële zekerheid is belangrijk voor mijn welzijn.
6	IMS6	One can only have a decent life with a lot of money.	Men kan alleen een fatsoenlijk leven leiden met veel geld.
7	IMS7	I enjoy material things.	Ik geniet van materiële dingen.
8	IMS8	To make more money, I would work more immediately.	Om meer geld te verdienen, zou ik direct meer gaan werken.

5-point Likert scale disagree strongly – agree strongly

Appendix E: Responses To Scale Items

		Study 1					Study 2		
Name	**Question**	N	Mean	*SD*	Min	Max	N	Mean	*SD*
cdif1	It was easy to choose between energy saving measures	202	1.332	1.317	-3	3	160	1.36	1.36
cdif2	I changed my mind several times while choosing energy saving measures	202	-0.51	1.603	-3	3	160	-0.54	1.57
cdif3	The task of choosing energy saving measures was overwhelming	202	-0.574	1.718	-3	3	160	-0.68	1.70
cdif4	Comparing the energy saving measures took a lot of effort	202	-0.96	1.574	-3	3	160	-1.11	1.53
pf1	I think it would take me little effort to perform the chosen measures.	202	0.342	1.692	-3	3	160	0.37	1.63
pf3	I think the chosen measures are easy to apply in my home environment.	202	0.807	1.554	-3	3	160	0.89	1.46
pf4	I think the chosen measures would be hard to perform.	202	-0.713	1.528	-3	3	160	-0.74	1.51
chsat1	I am satisfied with the measures I chose	202	1.644	0.942	-3	3	160	1.71	0.85
chsat2	I think I would enjoy performing the chosen energy saving measures	202	0.792	1.416	-3	3	160	0.86	1.35
chsat3	I would recommend the chosen measures to others	202	1.317	1.253	-3	3	160	1.41	1.17
chsat4	I will implement all the measures I have chosen	202	0.361	1.309	-3	3	160	0.48	1.26
chsat5	I think I chose the best energy saving measures from the list	202	-0.0149	1.594	-3	3	160	0.09	1.56
syssat1	I make better choices using the Saving aid.	202	1.163	1.261	-3	3	160	1.31	1.15
syssat2	The Saving aid is helpful to find appropriate measures.	202	1.594	1.019	-3	3	160	1.73	0.90
syssat3	I would recommend the Saving aid to others	202	1.406	1.239	-3	3	160	1.48	1.21
syssat4	I would use the Saving aid more often if possible	202	1.163	1.229	-3	3	160	1.23	1.17
syssat5	The Saving aid was useless	202	-1.95	1.188	-3	3	160	-2.08	1.07
gsup1	The Saving aid makes saving energy easier	202	1.351	1.129	-3	3	160	1.48	1.06
gsup2	The Saving aid motivates me to save more energy.	202	1.277	1.263	-3	3	160	1.41	1.16
gsup3	I think I will save more energy in the coming year thanks to the Saving aid	202	0.861	1.308	-3	3	160	1.01	1.22

gsup4	The Saving aid gives me more insight into the energy consumption of devices and systems in my home	202	1.574	1.326	-3	3	160	1.78	1.16
gsup5	The Saving aid made me more energy-conscious	202	1.238	1.358	-3	3	160	1.48	1.14
gsup6	The Saving aid makes me more aware of my options for saving energy	202	1.743	1.117	-3	3	160	1.90	1.01
sef1	I will be able to achieve most of the goals that I have set for myself concerning energy conservation.	202	1.084	1.28	-3	3	160	1.15	1.23
sef2	When facing difficult decisions on energy conservation, I am certain that I will accomplish them.	202	0.55	1.254	-3	3	160	0.66	1.17
sef3	In general, I think I can obtain energy conservation outcomes that are important to me.	202	1.178	1.141	-3	3	160	1.29	1.03
sef4	I believe I can succeed at most any energy conservation endeavour to which I set my mind.	202	1.074	1.277	-3	3	160	1.15	1.25
sef5	I am confident that I can perform effectively on many different tasks relating to energy conservation.	202	1.069	1.153	-3	3	160	1.13	1.09
sef6	Compared to other people, I can do most energy conservation task very well	202	0.366	1.387	-3	3	160	0.53	1.26
nep1	We are approaching the limit of the number of people the earth can support	202	0.901	1.32	-2	2	160	0.88	1.37
nep2	Humans have the right to modify the natural environment to suit their needs	202	-0.525	1.138	-2	2	160	-0.48	1.15
nep3	When humans interfere with nature it often produces disastrous consequences	202	0.921	0.989	-2	2	160	0.94	0.96
nep4	Human ingenuity will ensure that we do NOT make the earth unliveable	202	0.119	1.1	-2	2	160	0.19	1.07
nep5	Humans are severely abusing the environment	202	1.495	0.761	-1	2	160	1.53	0.70
nep6	The earth has plenty of natural resources if we just learn how to develop them	202	0.149	1.307	-2	2	160	0.14	1.32
nep7	Plants and animals have as much right as humans to exist	202	1.337	0.949	-2	2	160	1.33	0.91

nep8	The balance of nature is strong enough to cope with the impacts of modern industrial nations	202	-1.059	1.063	-2	2	160	-1.03	1.05
nep9	Despite our special abilities humans are still subject to the laws of nature	202	1.356	0.842	-2	2	160	1.39	0.77
nep10	The so-called “ecological crisis” facing humankind has been greatly exaggerated	202	-1.248	1.031	-2	2	160	-1.27	1.01
nep11	The earth is like a spaceship with very limited room and resources	202	0.738	1.081	-2	2	160	0.68	1.11
nep12	Humans were meant to rule over the rest of nature	202	-1.228	1.001	-2	2	160	-1.22	0.99
nep13	The balance of nature is very delicate and easily upset	202	0.748	1.056	-2	2	160	0.76	1.03
nep14	Humans will eventually learn enough about how nature works to be able to control it	202	-0.139	1.079	-2	2	160	-0.15	1.11
nep15	If things continue on their present course, we will soon experience a major ecological catastrophe	202	1.183	0.931	-2	2	160	1.19	0.91
ims1	Material wealth is important for me.	202	-0.0891	1.274	-2	2	160	-0.04	1.26
ims2	Money is important for me.	202	0.525	1.138	-2	2	160	0.58	1.08
ims3	Money makes me happy.	202	0.332	1.048	-2	2	160	0.39	1.06
ims4	Financial security is important for my well-being.	202	1.45	0.677	-1	2	160	1.52	0.62
ims6	One can only have a decent life with a lot of money.	202	-0.703	1.21	-2	2	160	-0.66	1.23
ims7	I enjoy material things.	202	0.307	1.053	-2	2	160	0.29	1.06
ims8	To make more money, I would work more immediately.	202	-0.119	1.236	-2	2	160	-0.01	1.24

Appendix F: Item-Level Data

ID	name	shown Current	Chosen Current	NA Current	Shown New	Chosen New	Removed New	Diffi-culty
1	Place brushes/valves on mailboxes and/or keyholes.	22	11	6	31	5	12	-0.7
2	Ensure good air circulation in your refrigerator.	34	14	3	24	8	4	-0.9
3	Throw away large food scraps, but don't pre-wash.	36	22	8	72	16	27	-0.3
4	Clean lamps and their sockets regularly.	29	5	0	17	6	6	1.81
5	Clean the coils at the back of your refrigerator regularly.	38	7	1	3	0	2	2.97
6	Regularly replace/clean the filters or bag of your vacuum cleaner.	22	14	1	100	38	43	0
7	Turn off the oven 10 minutes earlier.	30	8	1	7	0	5	2.63
8	Set boiler temperature to 60-65°C.	39	15	17	103	40	29	0.15
9	Save your laundry.	26	25	0	0	0	0	-3.6
10	Turn off the air conditioning in unused rooms.	29	14	14	22	2	13	-0.9
11	Turn off the ventilation in your bathroom for a maximum of 20 minutes after showering.	29	13	12	11	4	2	-1.2
12	Insulate your water heater so it feels cold to the touch.	31	3	12	58	17	17	0.8
13	Insulate your attic floor, including attic door or hatch.	25	4	13	31	5	15	1.25
14	Replace light bulbs with energy saving bulbs or better.	23	19	2	3	0	3	-2
15	Install windows with double glazing or better.	24	18	3	4	2	2	-1.6
16	Place light switches at both ends of a hallway.	34	19	4	11	1	5	-1.3

17	Use an energy-efficient monitor for your desktop computer.	25	10	4	93	25	35	0.37
18	Replace incandescent and energy-saving lamps with LED lamps.	23	19	1	4	2	1	-1.7
19	Buy a laptop or mini PC instead of a desktop computer.	28	20	1	1	0	0	-2.4
20	Do not boil more water than you need.	34	23	0	9	6	3	-1.4
21	Apply weatherstripping to doors.	33	12	4	97	41	21	0
22	Turn off your PC when you are not using it.	19	11	0	30	17	4	-0.6
23	Buy an energy-efficient washing machine.	25	8	8	66	10	21	-0.3
24	Install a shower water consumption meter.	27	3	2	3	3	1	2.89
25	Clean up leaves with a rake instead of a leaf blower.	40	20	20	4	0	4	-1.8
26	Install solar garden lights.	40	8	13	17	7	6	1.72
27	Thaw food in the refrigerator or sink, rather than in the microwave.	16	11	3	23	7	8	-1.1
28	Prune plants with a non-electric pruning or hedge trimmer	35	13	18	91	15	39	-0.1
29	Buy a hand lawnmower instead of a motorized lawnmower.	22	2	13	16	4	6	2.1
30	Set your TV to be energy efficient.	36	8	12	35	14	14	1.1
31	Insulate your cavity walls.	29	13	13	95	14	28	0
32	Let food cool before putting it in the fridge.	31	26	0	1	1	0	-2.3
33	Place rugs on your floor.	21	8	1	105	37	40	0.14
34	Wash at a low(er) temperature.	29	27	0	9	2	4	-1.4

35	Use a clothesline or drying rack instead of the dryer.	37	31	2	3	1	2	-2.2
36	Apply draft strips to the windows	27	10	5	30	16	6	-0.5
37	Seal cracks in exterior walls.	32	13	9	91	24	21	-0.1
38	Do not use a hairdryer to dry your hair.	35	18	4	24	5	6	-1
39	Place radiator foil (behind radiators).	32	8	8	17	7	8	1.96
40	Set your thermostat to a lower temperature if you are going away for a few days.	29	27	2	1	0	1	-2.3
41	Cook on an induction hob instead of a gas stove.	25	10	3	24	7	13	1.6
42	Install water-saving aerators on all your taps.	26	6	3	64	24	24	0.58
43	Shower short(er).	29	20	1	69	30	28	0.64
44	Turn off your water heater if you are going away for a few days.	29	6	9	35	9	16	1.01
45	Turn off the tap while you soap yourself.	25	14	0	33	12	16	1.16
46	Use a broom instead of a vacuum cleaner.	26	8	1	87	37	38	0.28
47	Repair leaky taps.	33	21	10	3	0	3	-2.1
48	Only run the dishwasher when it is full.	38	25	10	3	0	3	-2.2
49	Use a skillet instead of the oven.	28	20	1	22	7	8	-1.1
50	Cook with the correct size pan on the corresponding burner.	29	23	0	0	0	0	-2.7
51	Use a lid when cooking.	30	25	0	3	1	2	-2.2
52	Decide what you are going to take out of the fridge/freezer before you open it.	34	22	0	4	3	0	-1.7
53	Set your refrigerator to 4°C.	40	27	1	11	7	0	-1.2
54	Set your freezer to -18°C.	26	14	4	41	20	8	-0.4

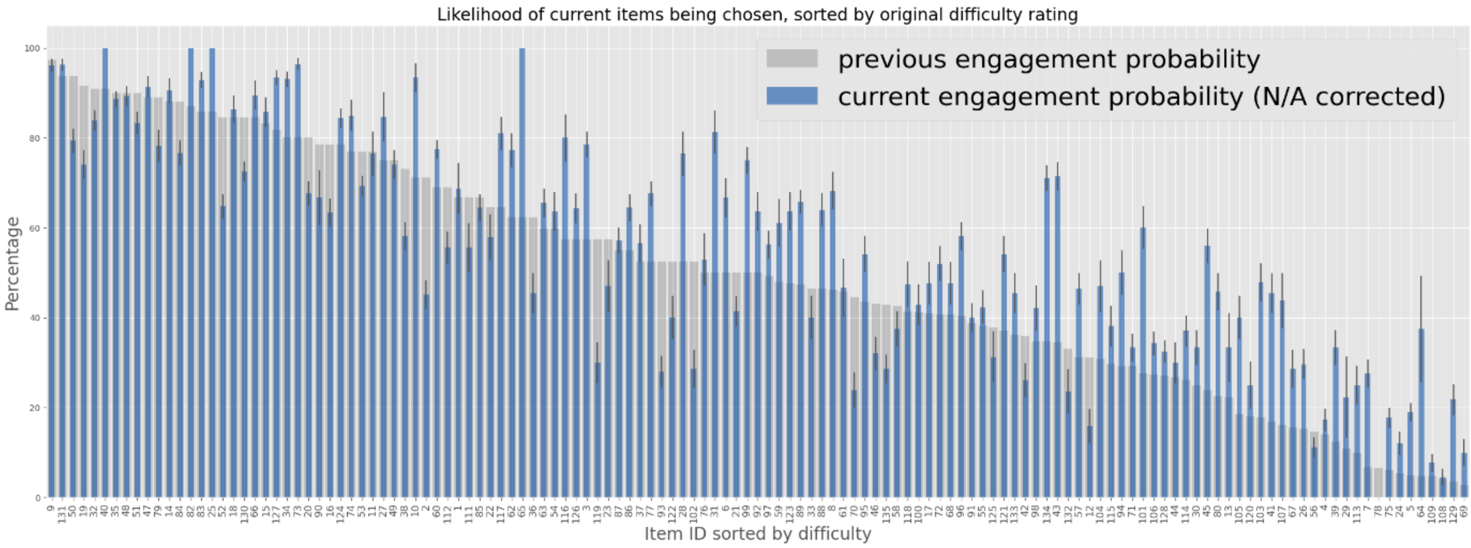
55	Install a programmable thermostat.	28	11	2	88	14	42	0.48
56	Install a remote thermostat.	33	3	6	17	9	5	1.77
57	Insulate ceilings.	39	13	11	56	7	32	0.79
58	Insulate heating pipes.	38	9	14	91	24	28	0.3
59	Insulate roofs.	33	11	15	105	16	33	0.08
60	Do not use an electric heater to heat large spaces.	45	31	5	29	5	18	-0.8
61	Seal holes in insulation material with polyurethane foam.	25	7	10	102	19	27	0.17
62	Lower the thermostat by 1°C.	23	17	1	30	6	14	-0.5
63	Set the thermostat to 14°C before going to sleep.	32	19	3	41	13	14	-0.4
64	Put timers on patio heaters.	36	3	28	3	1	0	2.99
65	Turn off the air conditioner when no one is home.	30	15	15	29	0	16	-0.5
66	Use an extra blanket instead of the heater.	20	17	1	4	1	2	-1.7
67	Use an electric blanket instead of the heater.	27	6	6	17	7	4	1.69
68	Buy an energy-efficient freezer.	28	10	7	90	30	39	0.38
69	Install a heat recovery system in the shower.	25	2	5	3	2	0	3.59
70	Install a water-saving shower head.	26	5	5	102	32	38	0.22
71	Install motion sensors on your lamps.	31	10	1	53	16	17	0.89
72	Install energy-efficient fixtures.	33	13	8	76	26	30	0.38
73	Use daylight and leave lights off as much as possible during the day.	28	27	0	9	3	5	-1.4
74	Dry only full loads in the dryer.	30	17	10	11	1	6	-1.2
75	Install an energy consumption manager.	40	6	6	6	2	3	2.72

76	Buy an energy-efficient dishwasher.	29	9	12	92	18	29	0
77	Enable power management of your desktop PC or laptop.	34	23	0	90	37	38	-0.1
78	Install a heat pump to heat the house.	29	0	6	7	2	0	2.66
79	Do not use an air conditioner.	24	18	1	3	1	2	-2.1
80	Renew your old water heater.	36	11	12	31	10	11	1.24
82	Take a shower instead of taking a bath.	30	27	3	3	0	3	-1.9
83	Do not use an electric blanket, but another blanket.	30	26	2	4	0	4	-1.8
84	Turn off the coffee maker completely.	34	23	4	3	2	1	-2
85	Air chambers 20 minutes daily.	33	20	2	31	15	12	-0.7
86	Wash by hand instead of in the dishwasher.	33	20	2	73	12	37	-0.2
87	Stir fry	39	20	4	73	26	24	-0.2
88	Make coffee without a hot plate.	34	16	9	102	17	46	0.14
89	Air clothes instead of washing them.	35	23	0	105	32	53	0.11
90	Turn off the dishwasher immediately after use.	22	10	7	11	1	5	-1.3
91	Turn off the washing machine completely.	33	12	3	85	33	32	0.46
92	Turn off the PC with switch box.	26	14	4	85	45	28	0
93	Buy an energy-efficient fridge/freezer combination.	30	7	5	88	28	20	-0.1
94	Keep extractor hood filter(s) clean.	24	10	4	51	14	30	0.89
95	Green power	32	13	8	100	28	37	0.26
96	Descale your coffee maker and/or kettle.	35	18	4	77	31	37	0.39
97	Free the fridge/freezer from ice.	34	18	2	108	58	38	0.03
98	Tumble dry shirts briefly instead of ironing them.	30	8	11	64	18	35	0.63

99	Carry out maintenance on the central heating boiler or geyser.	39	21	11	83	25	29	0
100	Day-night rate	29	9	8	87	20	38	0.36
101	Lower the boiler temperature.	26	12	6	50	12	26	0.97
102	Replace dimmers.	28	6	7	87	21	25	-0.1
103	Descale the washing machine.	26	11	3	23	8	13	1.53
104	Buy an energy-efficient dryer.	29	8	12	58	11	28	0.81
105	Install door closers.	31	8	11	24	8	9	1.48
106	Turn off the refrigerator when on vacation.	38	12	3	49	22	12	0.98
107	Install solar panels.	29	7	13	17	9	11	1.65
108	Wash with a hot-fill washing machine.	28	1	6	3	0	1	3.11
109	Install a solar water heater.	34	2	8	3	2	0	3.01
111	Iron several garments at a time.	34	10	16	30	3	8	-0.7
112	Use rechargeable batteries.	30	15	3	29	5	9	-0.8
113	Install a timer on your water heater.	30	5	10	15	3	4	2.21
114	Bleed your radiators regularly.	30	10	3	35	13	17	1.04
115	Use your tablet instead of your laptop/desktop.	24	8	3	58	14	25	0.86
116	Get rid of your second refrigerator.	27	12	12	60	9	24	-0.3
117	Use the eco program of your dishwasher.	38	17	17	30	9	12	-0.6
118	Install floor insulation.	27	9	8	88	18	23	0.35
119	Install thermostatic mixing valves.	24	6	4	59	13	13	-0.3
120	Install exterior wall insulation.	32	4	16	24	6	8	1.52
121	Make sure that the rubber seals of your refrigerator are kept clean and airtight.	25	13	1	88	39	26	0.53
122	Use an energy-efficient TV.	33	8	13	78	22	19	-0.1
123	Place your refrigerator in the right place.	22	14	0	104	17	55	0.1

124	Open and close curtains and the like at the right times.	35	27	3	11	4	4	-1.3
125	Place deciduous trees around your home.	28	5	12	88	20	18	0.5
126	Unplug devices and chargers when not in use.	29	18	1	59	31	15	-0.3
127	Turn off screens when not in use.	30	28	0	7	4	2	-1.5
128	Turn off your computer screen while downloading.	40	12	3	33	24	17	0.99
129	Cook in a pressure cooker.	24	5	1	3	0	2	3.32
130	Don't use a screen saver.	41	29	1	4	2	2	-1.7
131	Do not leave your extractor hood on for an unnecessarily long time.	32	26	5	0	0	0	-2.7
132	Install underfloor heating (if your house is well insulated).	23	4	6	65	11	24	0.7
133	Watch one hour less TV per day.	30	10	8	64	23	32	0.57
134	Check the pressure in your boiler.	34	22	3	64	15	32	0.63
135	Install a smaller cistern on your toilet(s).	39	8	11	90	26	24	0.29

Appendix G: Additional Figure Results



Appendix H: SEM Initial Study Important Output – mPlus

Green = Significant indirect effect

Blue = Almost significant indirect effect

Yellow = Indirect effect not significant

Orange = Scale or coefficient reversed (SEF and CDIF in initial study)

```
SUMMARY OF ANALYSIS

Number of groups                                                 1
Number of observations                                         202

Number of dependent variables                                   17
Number of independent variables                                  5
Number of continuous latent variables                            4

Observed dependent variables

  Continuous
   LOGKWH

  Binary and ordered categorical (ordinal)
   CDIF1       CDIF3       CDIF4       GSUP4       GSUP5       GSUP6
   SYSSAT2     CHSAT1      CHSAT2      CHSAT3      SEF1        SEF2
   SEF3        SEF4        SEF5        SEF6

Observed independent variables
   CO2         EUR         NEP         NEPXCO2     NEPXEUR

Continuous latent variables
   CDIF        GSUPS       CHSAT       SEF

MODEL FIT INFORMATION

Number of Free Parameters                        126

Chi-Square Test of Model Fit

          Value                              253.997*
          Degrees of Freedom                     193
          P-Value                             0.0021

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                             0.040
          90 Percent C.I.                      0.025  0.052
          Probability RMSEA <= .05             0.909

CFI/TLI
          CFI                                  0.981
          TLI                                  0.979

Chi-Square Test of Model Fit for the Baseline Model

          Value                             3479.390
          Degrees of Freedom                     221
          P-Value                             0.0000

SRMR (Standardized Root Mean Square Residual)
```

```
          Value                                  0.051

Optimum Function Value for Weighted Least-Squares Estimator

          Value                         0.51923866D+00


MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 CDIF     BY
    CDIF3             -0.529      0.057     -9.199      0.000
    CDIF1              0.730      0.062     11.769      0.000
    CDIF4             -0.683      0.054    -12.649      0.000

 GSUPS    BY
    SYSSAT2            0.736      0.042     17.540      0.000
    GSUP4              0.770      0.036     21.102      0.000
    GSUP5              0.774      0.036     21.342      0.000
    GSUP6              0.784      0.037     21.346      0.000

 CHSAT    BY
    CHSAT1             0.492      0.061      8.087      0.000
    CHSAT2             0.411      0.045      9.097      0.000
    CHSAT3             0.466      0.053      8.855      0.000

 SEF      BY
    SEF1               0.576      0.033     17.648      0.000
    SEF2               0.531      0.033     16.331      0.000
    SEF3               0.658      0.031     21.173      0.000
    SEF4               0.611      0.032     19.340      0.000
    SEF5               0.719      0.032     22.349      0.000
    SEF6               0.524      0.036     14.675      0.000

 CHSAT    ON
    CDIF               0.810      0.167      4.859      0.000
    GSUPS              0.570      0.136      4.201      0.000

 SEF      ON
    CHSAT              0.285      0.078      3.648      0.000
    CDIF               0.375      0.120      3.127      0.002

 GSUPS    ON
    CDIF               0.379      0.087      4.343      0.000

 CHSAT    ON
    LOGKWH             0.203      0.074      2.765      0.006

 CDIF     ON
    CO2               -1.424      0.392     -3.635      0.000
    NEP               -0.444      0.128     -3.473      0.001
    EUR               -0.985      0.430     -2.289      0.022
    NEPXEUR            0.426      0.197      2.166      0.030
    NEPXCO2            0.552      0.196      2.812      0.005

 LOGKWH   ON
    GSUPS              0.486      0.081      6.035      0.000

 Intercepts
    LOGKWH             8.073      0.510     15.836      0.000
```

```
R-SQUARE

    Observed                        Residual
    Variable         Estimate       Variance

    CDIF1               0.562          0.467
    CDIF3               0.303          0.720
    CDIF4               0.495          0.534
    GSUP4               0.681          0.323
    GSUP5               0.689          0.315
    GSUP6               0.706          0.297
    SYSSAT2             0.623          0.381
    CHSAT1              0.661          0.351
    CHSAT2              0.467          0.546
    CHSAT3              0.594          0.419
    SEF1                0.536          0.473
    SEF2                0.456          0.553
    SEF3                0.694          0.313
    SEF4                0.600          0.408
    SEF5                0.824          0.181
    SEF6                0.443          0.565
    LOGKWH              0.097

     Latent
    Variable         Estimate

    CDIF                0.109
    GSUPS               0.139
    CHSAT               0.645
    SEF                 0.391

TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS


                                                          Two-Tailed
                     Estimate       S.E.  Est./S.E.        P-Value

Effects from CO2 to LOGKWH

  Total                -0.262      0.096     -2.722          0.006
  Total indirect       -0.262      0.096     -2.722          0.006

  Specific indirect 1
    LOGKWH
    GSUPS
    CDIF
    CO2                -0.262      0.096     -2.722          0.006


Effects from EUR to LOGKWH

  Total                -0.181      0.093     -1.959          0.050
  Total indirect       -0.181      0.093     -1.959          0.050

  Specific indirect 1
    LOGKWH
    GSUPS
    CDIF
    EUR                -0.181      0.093     -1.959          0.050


Effects from NEP to LOGKWH

  Total                -0.082      0.030     -2.768          0.006
  Total indirect       -0.082      0.030     -2.768          0.006
```

```
  Specific indirect 1
    LOGKWH
    GSUPS
    CDIF
    NEP                 -0.082      0.030     -2.768      0.006

Effects from NEPXCO2 to LOGKWH

  Total                  0.102      0.045      2.279      0.023
  Total indirect         0.102      0.045      2.279      0.023

  Specific indirect 1
    LOGKWH
    GSUPS
    CDIF
    NEPXCO2              0.102      0.045      2.279      0.023

Effects from NEPXEUR to LOGKWH

  Total                  0.079      0.043      1.847      0.065
  Total indirect         0.079      0.043      1.847      0.065

  Specific indirect 1
    LOGKWH
    GSUPS
    CDIF
    NEPXEUR              0.079      0.043      1.847      0.065


Effects from CO2 to SEF

  Total                 -0.965      0.292     -3.306      0.001
  Total indirect        -0.965      0.292     -3.306      0.001

  Specific indirect 1
    SEF
    CDIF
    CO2                 -0.534      0.225     -2.376      0.017

  Specific indirect 2
    SEF
    CHSAT
    CDIF
    CO2                 -0.329      0.119     -2.756      0.006

  Specific indirect 3
    SEF
    CHSAT
    GSUPS
    CDIF
    CO2                 -0.088      0.038     -2.326      0.020

  Specific indirect 4
    SEF
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    CO2                 -0.015      0.008     -1.985      0.047
```

```
Effects from EUR to SEF

  Total                  -0.668       0.311     -2.145       0.032
  Total indirect         -0.668       0.311     -2.145       0.032

  Specific indirect 1
    SEF
    CDIF
    EUR                  -0.369       0.208     -1.778       0.075

  Specific indirect 2
    SEF
    CHSAT
    CDIF
    EUR                  -0.227       0.110     -2.061       0.039

  Specific indirect 3
    SEF
    CHSAT
    GSUPS
    CDIF
    EUR                  -0.061       0.034     -1.773       0.076

  Specific indirect 4
    SEF
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    EUR                  -0.011       0.006     -1.700       0.089


Effects from NEP to SEF

  Total                  -0.301       0.096     -3.139       0.002
  Total indirect         -0.301       0.096     -3.139       0.002

  Specific indirect 1
    SEF
    CDIF
    NEP                  -0.166       0.073     -2.274       0.023

  Specific indirect 2
    SEF
    CHSAT
    CDIF
    NEP                  -0.102       0.038     -2.715       0.007

  Specific indirect 3
    SEF
    CHSAT
    GSUPS
    CDIF
    NEP                  -0.027       0.012     -2.347       0.019

  Specific indirect 4
    SEF
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    NEP                  -0.005       0.002     -2.001       0.045
```

```
Effects from NEPXCO2 to SEF

  Total                 0.374      0.142      2.630      0.009
  Total indirect        0.374      0.142      2.630      0.009

  Specific indirect 1
    SEF
    CDIF
    NEPXCO2             0.207      0.100      2.071      0.038

  Specific indirect 2
    SEF
    CHSAT
    CDIF
    NEPXCO2             0.127      0.054      2.362      0.018

  Specific indirect 3
    SEF
    CHSAT
    GSUPS
    CDIF
    NEPXCO2             0.034      0.017      2.043      0.041

  Specific indirect 4
    SEF
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    NEPXCO2             0.006      0.003      1.785      0.074


Effects from NEPXEUR to SEF

  Total                 0.289      0.142      2.041      0.041
  Total indirect        0.289      0.142      2.041      0.041

  Specific indirect 1
    SEF
    CDIF
    NEPXEUR             0.160      0.093      1.719      0.086

  Specific indirect 2
    SEF
    CHSAT
    CDIF
    NEPXEUR             0.098      0.050      1.971      0.049

  Specific indirect 3
    SEF
    CHSAT
    GSUPS
    CDIF
    NEPXEUR             0.026      0.015      1.709      0.087

  Specific indirect 4
    SEF
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    NEPXEUR             0.005      0.003      1.601      0.109
```

```
Effects from CO2 to CHSAT

  Total                 -1.513       0.500     -3.030       0.002
  Total indirect        -1.513       0.500     -3.030       0.002

  Specific indirect 1
    CHSAT
    CDIF
    CO2                 -1.153       0.404     -2.851       0.004

  Specific indirect 2
    CHSAT
    GSUPS
    CDIF
    CO2                 -0.308       0.116     -2.662       0.008

  Specific indirect 3
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    CO2                 -0.053       0.028     -1.891       0.059


Effects from EUR to CHSAT

  Total                 -1.047       0.504     -2.079       0.038
  Total indirect        -1.047       0.504     -2.079       0.038

  Specific indirect 1
    CHSAT
    CDIF
    EUR                 -0.797       0.391     -2.042       0.041

  Specific indirect 2
    CHSAT
    GSUPS
    CDIF
    EUR                 -0.213       0.114     -1.865       0.062

  Specific indirect 3
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    EUR                 -0.037       0.023     -1.611       0.107


Effects from NEP to CHSAT

  Total                 -0.472       0.162     -2.920       0.003
  Total indirect        -0.472       0.162     -2.920       0.003

  Specific indirect 1
    CHSAT
    CDIF
    NEP                 -0.359       0.131     -2.741       0.006

  Specific indirect 2
    CHSAT
    GSUPS
    CDIF
    NEP                 -0.096       0.036     -2.636       0.008

  Specific indirect 3
    CHSAT
```

```
    LOGKWH
    GSUPS
    CDIF
    NEP                  -0.017       0.009      -1.884       0.060


Effects from NEPXCO2 to CHSAT

  Total                   0.587       0.237       2.478       0.013
  Total indirect          0.587       0.237       2.478       0.013

  Specific indirect 1
    CHSAT
    CDIF
    NEPXCO2               0.447       0.187       2.390       0.017

  Specific indirect 2
    CHSAT
    GSUPS
    CDIF
    NEPXCO2               0.119       0.053       2.235       0.025

  Specific indirect 3
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    NEPXCO2               0.021       0.012       1.704       0.088


Effects from NEPXEUR to CHSAT

  Total                   0.453       0.228       1.986       0.047
  Total indirect          0.453       0.228       1.986       0.047

  Specific indirect 1
    CHSAT
    CDIF
    NEPXEUR               0.345       0.176       1.958       0.050

  Specific indirect 2
    CHSAT
    GSUPS
    CDIF
    NEPXEUR               0.092       0.051       1.794       0.073

  Specific indirect 3
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    NEPXEUR               0.016       0.010       1.528       0.126
```

Appendix I: Additional Tables And Figures Follow-Up Study

Table H.1

Housing information follow-up study.

Energy label	Freq.	Percent	House type	Freq.	Percent
A/A+	19	11.88	apartment	57	35.63
B	13	8.13			
C	20	12.50	(partly) freestanding House	38	23.75
D	7	4.38			
E	5	3.13	room	25	15.63
F	4	2.50			
G/G-	7	4.38	Terraced House	39	24.38
Prefer Not to say	85	53.13			
			other	1	0.63
Total	160	100.00	Total	160	100.00

Appendix J: SEM Follow-Up Study Important Output – mPlus

Green = Significant indirect effect

Blue = Almost significant indirect effect

Yellow = Indirect effect not significant

Orange = Scale or coefficient reversed (CDIF in follow-up study)

```
SUMMARY OF ANALYSIS

Number of groups                                                 1
Number of observations                                         160

Number of dependent variables                                   17
Number of independent variables                                  5
Number of continuous latent variables                            4

Observed dependent variables

  Continuous
   LOGKWH

  Binary and ordered categorical (ordinal)
   CDIF1       CDIF3       CDIF4       GSUP4       GSUP5       GSUP6
   SYSSAT2     CHSAT1      CHSAT2      CHSAT3      SEF1        SEF2
   SEF3        SEF4        SEF5        SEF6

Observed independent variables
   CO2         EUR         NEP         NEPXCO2     NEPXEUR

Continuous latent variables
   CDIF        GSUPS       CHSAT       SEF

Chi-Square Test of Model Fit

          Value                            252.054*
          Degrees of Freedom                   193
          P-Value                           0.0027

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.044
          90 Percent C.I.                    0.027  0.058
          Probability RMSEA <= .05           0.749

CFI/TLI

          CFI                                0.971
          TLI                                0.967

Chi-Square Test of Model Fit for the Baseline Model

          Value                           2277.352
          Degrees of Freedom                   221
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                              0.065

Optimum Function Value for Weighted Least-Squares Estimator

          Value                    0.68473157D+00
```

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
CDIF BY				
CDIF3	-0.504	0.058	-8.747	0.000
CDIF1	0.826	0.067	12.344	0.000
CDIF4	-0.630	0.055	-11.531	0.000
GSUPS BY				
SYSSAT2	0.697	0.050	14.031	0.000
GSUP4	0.777	0.047	16.482	0.000
GSUP5	0.775	0.043	18.003	0.000
GSUP6	0.827	0.040	20.454	0.000
CHSAT BY				
CHSAT1	0.509	0.075	6.795	0.000
CHSAT2	0.387	0.048	7.983	0.000
CHSAT3	0.490	0.068	7.238	0.000
SEF BY				
SEF1	0.572	0.037	15.618	0.000
SEF2	0.585	0.041	14.414	0.000
SEF3	0.710	0.035	20.377	0.000
SEF4	0.651	0.036	18.008	0.000
SEF5	0.731	0.037	19.826	0.000
SEF6	0.580	0.039	14.679	0.000
CHSAT ON				
CDIF	1.004	0.206	4.882	0.000
GSUPS	0.556	0.149	3.728	0.000
SEF ON				
CHSAT	0.245	0.082	2.982	0.003
CDIF	0.245	0.128	1.914	0.056
GSUPS ON				
CDIF	0.217	0.085	2.542	0.011
CHSAT ON				
LOGKWH	0.180	0.069	2.592	0.010
CDIF ON				
CO2	-1.565	0.461	-3.397	0.001
NEP	-0.507	0.183	-2.777	0.005
NEPXCO2	0.690	0.253	2.722	0.006
EUR	-0.507	0.555	-0.914	0.361
NEPXEUR	0.333	0.260	1.280	0.200
LOGKWH ON				
GSUPS	0.707	0.193	3.669	0.000
Intercepts				
LOGKWH	5.441	0.817	6.659	0.000

R-SQUARE

Observed Variable	Estimate	Residual Variance
CDIF1	0.714	0.317
CDIF3	0.283	0.746
CDIF4	0.432	0.603
GSUP4	0.633	0.369
GSUP5	0.631	0.371
GSUP6	0.717	0.285
SYSSAT2	0.510	0.492
CHSAT1	0.779	0.233
CHSAT2	0.460	0.557
CHSAT3	0.723	0.291
SEF1	0.458	0.550
SEF2	0.479	0.529

```
    SEF3               0.701      0.306
    SEF4               0.591      0.417
    SEF5               0.742      0.264
    SEF6               0.470      0.537
    LOGKWH             0.095

     Latent
    Variable        Estimate

    CDIF               0.136
    GSUPS              0.052
    CHSAT              0.684
    SEF                0.296

TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Effects from CO2 to LOGKWH

  Total               -0.240      0.132     -1.815      0.070
  Total indirect      -0.240      0.132     -1.815      0.070

  Specific indirect 1
    LOGKWH
    GSUPS
    CDIF
    CO2               -0.240      0.132     -1.815      0.070


Effects from EUR to LOGKWH

  Total               -0.078      0.098     -0.798      0.425
  Total indirect      -0.078      0.098     -0.798      0.425

  Specific indirect 1
    LOGKWH
    GSUPS
    CDIF
    EUR               -0.078      0.098     -0.798      0.425


Effects from NEP to LOGKWH

  Total               -0.078      0.044     -1.751      0.080
  Total indirect      -0.078      0.044     -1.751      0.080

  Specific indirect 1
    LOGKWH
    GSUPS
    CDIF
    NEP               -0.078      0.044     -1.751      0.080


Effects from NEPXCO2 to LOGKWH

  Total                0.106      0.062      1.698      0.089
  Total indirect       0.106      0.062      1.698      0.089

  Specific indirect 1
    LOGKWH
    GSUPS
    CDIF
    NEPXCO2            0.106      0.062      1.698      0.089


Effects from NEPXEUR to LOGKWH

  Total                0.051      0.049      1.042      0.298
  Total indirect       0.051      0.049      1.042      0.298

  Specific indirect 1
```

```
    LOGKWH
    GSUPS
    CDIF
    NEPXEUR              0.051       0.049       1.042       0.298


Effects from CO2 to SEF

  Total                 -0.826       0.273      -3.025       0.002
  Total indirect        -0.826       0.273      -3.025       0.002

  Specific indirect 1
    SEF
    CDIF
    CO2                 -0.384       0.224      -1.709       0.087

  Specific indirect 2
    SEF
    CHSAT
    CDIF
    CO2                 -0.386       0.164      -2.358       0.018

  Specific indirect 3
    SEF
    CHSAT
    GSUPS
    CDIF
    CO2                 -0.046       0.028      -1.670       0.095

  Specific indirect 4
    SEF
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    CO2                 -0.011       0.007      -1.466       0.143


Effects from EUR to SEF

  Total                 -0.268       0.299      -0.897       0.370
  Total indirect        -0.268       0.299      -0.897       0.370

  Specific indirect 1
    SEF
    CDIF
    EUR                 -0.124       0.153      -0.815       0.415

  Specific indirect 2
    SEF
    CHSAT
    CDIF
    EUR                 -0.125       0.141      -0.888       0.374

  Specific indirect 3
    SEF
    CHSAT
    GSUPS
    CDIF
    EUR                 -0.015       0.019      -0.783       0.434

  Specific indirect 4
    SEF
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    EUR                 -0.003       0.004      -0.784       0.433

Effects from NEP to SEF

  Total                 -0.268       0.107      -2.501       0.012
  Total indirect        -0.268       0.107      -2.501       0.012
```

```
  Specific indirect 1
    SEF
    CDIF
    NEP                -0.124      0.079     -1.580      0.114

  Specific indirect 2
    SEF
    CHSAT
    CDIF
    NEP                -0.125      0.059     -2.108      0.035

  Specific indirect 3
    SEF
    CHSAT
    GSUPS
    CDIF
    NEP                -0.015      0.009     -1.582      0.114

  Specific indirect 4
    SEF
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    NEP                -0.003      0.002     -1.423      0.155


Effects from NEPXCO2 to SEF

  Total                 0.364      0.148      2.466      0.014
  Total indirect        0.364      0.148      2.466      0.014

  Specific indirect 1
    SEF
    CDIF
    NEPXCO2             0.169      0.108      1.569      0.117

  Specific indirect 2
    SEF
    CHSAT
    CDIF
    NEPXCO2             0.170      0.081      2.104      0.035

  Specific indirect 3
    SEF
    CHSAT
    GSUPS
    CDIF
    NEPXCO2             0.020      0.013      1.539      0.124

  Specific indirect 4
    SEF
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    NEPXCO2             0.005      0.003      1.405      0.160


Effects from NEPXEUR to SEF

  Total                 0.176      0.142      1.241      0.215
  Total indirect        0.176      0.142      1.241      0.215

  Specific indirect 1
    SEF
    CDIF
    NEPXEUR             0.082      0.077      1.062      0.288

  Specific indirect 2
    SEF
    CHSAT
    CDIF
```

```
    NEPXEUR             0.082       0.069       1.194       0.232

  Specific indirect 3
    SEF
    CHSAT
    GSUPS
    CDIF
    NEPXEUR             0.010       0.010       0.998       0.318

  Specific indirect 4
    SEF
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    NEPXEUR             0.002       0.002       0.969       0.332


Effects from CO2 to CHSAT

  Total                -1.804       0.654      -2.760       0.006
  Total indirect       -1.804       0.654      -2.760       0.006

  Specific indirect 1
    CHSAT
    CDIF
    CO2                -1.572       0.581      -2.706       0.007

  Specific indirect 2
    CHSAT
    GSUPS
    CDIF
    CO2                -0.189       0.098      -1.935       0.053

  Specific indirect 3
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    CO2                -0.043       0.028      -1.540       0.124


Effects from EUR to CHSAT

  Total                -0.584       0.661      -0.884       0.377
  Total indirect       -0.584       0.661      -0.884       0.377

  Specific indirect 1
    CHSAT
    CDIF
    EUR                -0.509       0.572      -0.890       0.374

  Specific indirect 2
    CHSAT
    GSUPS
    CDIF
    EUR                -0.061       0.077      -0.796       0.426

  Specific indirect 3
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    EUR                -0.014       0.018      -0.786       0.432


Effects from NEP to CHSAT

  Total                -0.584       0.248      -2.352       0.019
  Total indirect       -0.584       0.248      -2.352       0.019

  Specific indirect 1
    CHSAT
    CDIF
```

```
    NEP                -0.509      0.220     -2.313      0.021

  Specific indirect 2
    CHSAT
    GSUPS
    CDIF
    NEP                -0.061      0.034     -1.788      0.074

  Specific indirect 3
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    NEP                -0.014      0.009     -1.482      0.138


Effects from NEPXCO2 to CHSAT

  Total                 0.795      0.336      2.365      0.018
  Total indirect        0.795      0.336      2.365      0.018

  Specific indirect 1
    CHSAT
    CDIF
    NEPXCO2             0.692      0.296      2.342      0.019

  Specific indirect 2
    CHSAT
    GSUPS
    CDIF
    NEPXCO2             0.083      0.048      1.740      0.082

  Specific indirect 3
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    NEPXCO2             0.019      0.013      1.470      0.142


Effects from NEPXEUR to CHSAT

  Total                 0.384      0.316      1.217      0.223
  Total indirect        0.384      0.316      1.217      0.223

  Specific indirect 1
    CHSAT
    CDIF
    NEPXEUR             0.335      0.273      1.226      0.220

  Specific indirect 2
    CHSAT
    GSUPS
    CDIF
    NEPXEUR             0.040      0.039      1.043      0.297

  Specific indirect 3
    CHSAT
    LOGKWH
    GSUPS
    CDIF
    NEPXEUR             0.009      0.009      0.986      0.324
```