UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Goal setting and signposting in a rasch-based recommender system to promote household energy conservation

Oonk, Aleid S.; Starke, Alain D.; Willemsen, Martijn C.

DOI
10.1007/s12053-025-10332-2

Publication date
2025

Document Version
Final published version

Published in
Energy Efficiency

License
CC BY

Link to publication

Citation for published version (APA):
Oonk, A. S., Starke, A. D., & Willemsen, M. C. (2025). Goal setting and signposting in a rasch-based recommender system to promote household energy conservation. *Energy Efficiency*, *18*(5), Article 48. https://doi.org/10.1007/s12053-025-10332-2

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:14 Oct 2026

Energy Efficiency (2025) 18:48
https://doi.org/10.1007/s12053-025-10332-2

ORIGINAL ARTICLE

Goal setting and signposting in a rasch-based recommender system to promote household energy conservation

Aleid S. Oonk · Alain D. Starke · Martijn C. Willemsen

Received: 14 May 2024 / Accepted: 1 May 2025 / Published online: 7 June 2025
© The Author(s) 2025

Abstract Recent studies have used algorithm tailoring on digital platforms to provide household energy-saving advice. Such 'recommender systems' have successfully used the psychometric Rasch model as an advice algorithm, matching energy-saving measures in terms of their difficulty to consumers' ability levels. While these previous studies indicated positive user experiences, tailored advice did not lead to higher savings overall; not even when also using persuasive nudges, such as displaying social norm percentages in the system. One possible reason for these results was that the system was used exploratively, allowing users to pick energy measures as they liked without tapping into goal setting or value-based motivational frames (e.g., signposts). In this study, 202 participants used and evaluated our 'Saving Aid' Rasch recommender system, choosing energy-saving measures they would like to perform at home. Through a 3×2-between subject design, we examined whether guided goal setting and signposts (kWh/Euro/CO2) affected user experience and energy savings. Following the signpost literature, we examined the moderation of these effects by user values, such as environmental concern (New Environmental Paradigm (NEP) score). A structural equation model analysis revealed that goal setting did not affect outcome variables, while signpost framing had varying effects, although these were not in line with prior expectations. Still, the overall system remains promising, with users achieving a 316 kWh yearly savings with the chosen recommendations.

Keywords Recommender systems · Sustainability · Energy saving · Goal setting · Signposting · Rasch model

A. S. Oonk · M. C. Willemsen (✉)
Human-Technology Interaction Group, Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Atlas 9.401, Box 513, 5600 MB Eindhoven, Netherlands
e-mail: M.C.Willemsen@tue.nl

A. S. Oonk
e-mail: aleidoonk@gmail.com

A. D. Starke
Persuasive Communication Group, Amsterdam School of Communication Research, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, Netherlands
e-mail: A.D.Starke@uva.nl

A. D. Starke
Department of Information Science and Media Studies, University of Bergen, Fosswinckels gate 6, 5007 Bergen, Norway

M. C. Willemsen
Jheronimus Academy of Data Science, St. Janssingel 92, 's, Hertogenbosch, Netherlands

Introduction

Efforts to reduce greenhouse gas emissions can be focused on the reduction of both industry and household energy usage. For the latter, individuals play a significant role, but they might defer from acting

due to the large number of options available to limit their energy usage (Gardner & Stern, 2008). It can be unclear which measures are the most effective for a specific individual (Attari & Rajagopal, 2015; Starke et al., 2020), making energy-efficient choices and behaviour a difficult task. While some measures often promoted by governments (e.g., solar PV, heat pumps, insulation) could drastically reduce one's energy consumption, they are often regarded to be too costly, financially, cognitively or in terms of time (Abrahamse et al., 2005; Boudet et al., 2016; Starke et al., 2020). On the other hand, many energy-saving tips promoted on websites and by utility companies are often already performed by most people or do not result in substantial savings (e.g., putting a lid on the pan while cooking).

Energy recommender systems are digital technologies that can help consumers to overcome choice overload issues, by finding energy-saving measures that suit their preferences (Knijnenburg et al., 2014). They use tailoring or personalization algorithms that present content to users that fit their preferences or needs (Jannach et al., 2010; Knijnenburg et al., 2014).

A recent example of an energy recommender system is the 'Saving Aid', developed by Starke et al. (2020). Contrary to most data-driven approaches in recommender systems work (Jannach et al., 2010), psychological theory informs the design of the algorithm (Starke et al., 2020). The psychometric Rasch model is employed to provide tailored household energy-saving advice to its users, assessing the behavioural difficulty of energy-saving measures and matching these to individuals based on their energy-saving ability. Whereas this earlier work has largely focused on validating this psychometric scale and introducing a social norm nudge (Starke et al., 2021), it has not addressed key mechanisms that relate to people's understanding and behaviour of energy-saving measures. Among them, the use of self-set goals and how information is presented (i.e., framing) are important behavioural determinants of conservation behaviour (Abrahamse et al., 2007; Ungemach et al., 2018). For example, saving metrics can be presented as kWh, Euro or CO2 reductions. In a study involving a car-selection task, it was found that such different attribute translations (signposts), may activate different objectives for users with different values (Ungemach et al., 2018). Additionally, goal setting has previously been shown to be effective in helping people achieve better outcomes (Locke & Latham, 2002), and more specifically, goal setting was found to be effective in reducing energy usage (Abrahamse et al., 2007; Harding & Hsiaw, 2014). In this paper, we explore the effects of goal setting and signposting within a Rasch-based energy-saving recommender system.

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), and are used to make sense of large quantities of data and a multitude of options (Lü et al., 2012). They are employed in various domains ranging from music (e.g., Spotify), movies, and series (Netflix), to travel, health, nutrition, and fitness applications. These systems can employ a one-size-fits-all approach (i.e., showing the most popular content overall), or personalize recommendations based on user characteristics, preferences, and/or behaviour. Recommender systems can suggest similar items based on item characteristics, so-called content-based filtering, or recommend items based on preferences of similar users, so-called collaborative filtering, (e.g., Koren et al., 2021), or a combination of both. While they provide many opportunities, they are also accompanied by various challenges; for example, tailored systems need a certain amount of user data to draw from, referred to as the cold start problem (Lika et al., 2014), preferences of users might change over time, and trade-offs between diversity and accuracy must be made to optimize these systems (Lü et al., 2012). Additionally, user interface characteristics might influence their effectiveness (Lü et al., 2012), and different users might respond differently to different systems (Knijnenburg et al., 2011). Small changes in these interfaces, such as user goals and different attribute translations that we examine in this paper, can change system outcomes.

Energy recommender systems

Digital technologies to encourage energy-saving behaviour are typically embedded in theories of persuasion (Adaji & Adisa, 2022). To this end, various

persuasive methods have been employed (Schultz, 2014; Warren et al., 2017). For example, one-size-fits-all messages or campaigns might highlight benefits to the environment or financial benefits of reducing energy usage or other sustainable behaviours. Doing so might increase effectiveness, for example when matched to personal values (Van den Broek et al., 2017). However, these persuasive messages are typically not tailored to the current behaviour or past preferences of an individual recipient.

Recommender systems in the energy-saving domain are generally based on the premise that behaviour can be altered, often through persuasive methods, e.g., through reward or cue-based systems aimed at habit changes (Alsalemi et al., 2019). From social judgment theory, we know that persuasive messages are most effective if they are within the so-called '*latitude of acceptance*'; a certain range of 'attitudes' close enough to one's own that someone is likely to be receptive to them (Eagly & Telaak, 1972). If persuasive messages convey attitudes that are too far removed from one's current stance, they are regarded as being in the '*latitude of rejection*' and are less likely to be effective (Eagly & Telaak, 1972). The 'widths' of these ranges differ per person, as do current attitudes, and therefore, energy-saving recommendations might also need to differ per person to be effective. In order to recommend someone with suitable energy-saving measures, it is therefore important to know what someone's current attitude towards energy-saving is.

One way of determining this is simply looking at the number of energy-saving behaviours someone has already adopted. This is consistent with methods used in recommender system research, which tap into historical data from its users (Jannach et al., 2010). Someone who already performs a large number of measures is considered to have a positive attitude towards energy saving. The psychometric Rasch model is based on this premise, and more specifically, based on the idea that energy-saving behaviour falls on a spectrum of low to high difficulty measures, which is in turn tied to low and high energy-saving ability levels.

Rasch model

The use of Rasch is embedded in an attitude paradigm called 'Campbell's Paradigm' (Kaiser et al., 2010), which aims to mitigate the attitude-behaviour gap. It assumes attitude and behaviour are two sides of the same coin (Kaiser et al., 2010), proposing a stochastic, 'logit' relation between individuals' attitudes and measures' difficulty when predicting whether a measure will be performed. An attitude becomes apparent by the increasingly difficult behavioural steps that an individual takes towards attaining an attitudinal goal, such as energy conservation. This is modelled on a one-dimensional scale that captures both persons and measures on their ability and difficulty (Kaiser et al., 2010; Starke & Willemsen, 2024). The scale assumes measures are more difficult if they are performed by fewer people, while a person who performs more measures is considered more able, having a stronger ability or attitude. The probability P that an individual n performs a certain measure i can be calculated using the following equation:

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

In which θ is the individual's ability, and δ is the measure's behavioural costs. This formula results in item-characteristic curves (ICC's) that depend on item difficulty, as can be seen in Fig. 1, indicating the likelihood (Y-axis) that an individual with a certain ability/attitude (X-axis), performs a certain item (different curves). When ability and difficulty match, the probability of engagement is 50%.

Because Rasch allows for sorting people and measures on the same one-dimensional scale, this approach seems especially suitable for tailored energy-saving advice and recommender systems; after a user's ability is determined, they can now be given meaningful

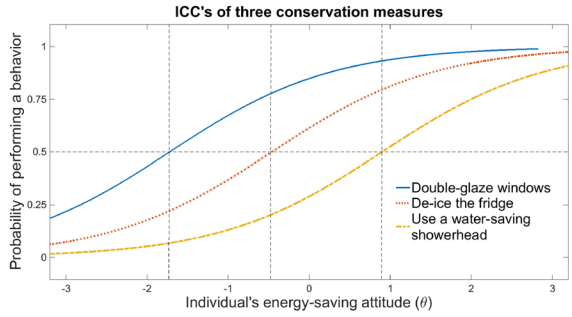


Fig. 1 Item characteristic curves of three energy-saving measures, obtained from Starke et al. (2020)

recommendations to curb their energy usage. For tailored advice, the Rasch model helps to explore the trade-off between a recommended measure's novelty and feasibility for a particular user (Starke et al., 2020).

Using the Rasch model as a tailoring algorithm, a recommender system was found to be more effective than a non-personalized system, in terms of user evaluations (Starke et al., 2020). However, it did not necessarily make users choose the measures with higher energy savings. Additional studies that aimed to support higher energy savings through nudges, e.g., by presenting 'algorithm fit scores', social norms, and smart saving scores (Bams, 2018; Starke et al., 2017, 2021), affected *which* measures were chosen, but did not lead to higher savings. Given that we can influence decisions in such a system, the next step is to determine how we can help people choose more efficient recommendations, such that overall savings do increase. One hypothesized problem is that kWh savings and their magnitude (e.g., how much is 100 kWh) are unclear. Attari et al. (2010) show that people tend to underestimate the effectiveness of conservation measures; participants in their study thought that curtailment measures (e.g., turn off the lights after leaving a room) were most effective for saving energy, as opposed to efficiency measures (e.g., insulate a cavity wall), which in reality tend to be higher in kWh savings. In line with these inaccurate expectations of kWh savings for household energy-saving measures, it might also be difficult for people to accurately judge what a reasonable amount of savings would be when using the recommender system, with respect to choosing a limited number of measures. Moreover, users in earlier Rasch recommender studies by Starke et al. (2020), might not have been committed or have had a specific goal in mind and thus opted for only few measures.

To mitigate the issues from earlier studies and to encourage higher savings in this study, we examine two different mechanisms in our novel energy recommender system. First, we investigate the effectiveness of including goal-setting functionality in our recommender system, which aims to mitigate erroneous kWh savings perceptions. In addition, we investigate value-based motivational frames, through signposting, by expressing environmental benefits in terms of other attributes. Signposting refers to attribute translations (e.g., kWh, Euro) in line with user values, with the aim of activating certain user objectives (Ungemach et al., 2018). We will discuss these aspects in more detail below. In doing so, we address the following overall research question:

- **[RQ]:** "What are the effects of goal setting and signposting on the user experience and energy savings in a Rasch-based energy recommender system?"

Goal setting

According to Locke and Latham (2002), the mere presence of goals increases performance, and an adequate balance between goal difficulty and feasibility leads to higher satisfaction. A goal that is assigned by an authority figure will furthermore strengthen the belief that this goal is reachable (Locke & Latham, 1990). At the same time, feeling personally responsible for reaching a goal, leads to higher feelings of competence (Eccles & Wigfield, 2002). Goals that are autonomously set furthermore lead to higher achievement (Koestner et al., 2008). Therefore, incorporating goal setting into an energy-saving recommender system might help users to achieve higher savings than they otherwise would.

Goal setting has shown promising, yet varying effects on energy use. Abrahamse et al. (2007), who consider direct energy usage (e.g., not energy use through travel), found a significant effect of goal setting and education on energy usage. Instructing households to reduce their energy consumption by 5%, along with performance feedback and education, resulted in an energy reduction of 5.1% on average over the course of a 5-month study. In a similar vein, Harding and Hsiaw (2014) asked their users to select energy-saving measures from a list of recommendations. They found that users who committed to reduce their current energy use between 0 and 15%, achieved the highest energy reductions at around 11%. Users who either chose no items, or whose item-choice resulted in overly optimistic goals, only saved between 1 and 1.5% on average. Interestingly, many users tended to over-commit, with 40% of them opting for an overly optimistic goal between 15 and 50%, and 12% of users opting for an even higher goal.

The previous research indicates that goal setting can affect the adoption of energy-saving measures. The goal-setting theory furthermore states

that self-chosen goals ensure a degree of autonomy. However, some direction might be needed to inform users what would be a reasonable goal for them. Following this, we presented users in the goal condition with three saving goals (an easy, moderate, and difficult goal) such that they had adequate guidance yet still had a degree of autonomy. We expected that this would lead to higher savings and higher system satisfaction than having no goals:

- **[H1]:** Guided goal setting leads to higher savings than having no goals

Signposting

Energy-saving metrics are not restricted to kWh. They can be presented in different units, such as CO2 emission reductions and monetary savings. The understanding and expected consequences of measures can depend on how the presented information is framed. 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. However, they did not distinguish between different kinds of customers who might be affected differently by different labels.

In a study by Ungemach et al. (2018), it was observed that people with stronger environmental concern, as measured by the NEP scale (New Environmental Paradigm), were more likely to pick an energy-efficient car when presented with information on its CO2 emissions per mile, as compared to its fuel usage per mile. Thus, the same exact information being presented with different units, resulted in different behaviour for different users. As a user could obtain the presented information by a simple scale translation, there seems to be no immediate reason for this effect. However, according to Ungemach et al. (2018), such attribute translations might activate different objectives for users with different personal values. This mechanism is referred to as 'signposting'; the presentation of information in line with user values, helps to guide users towards their personal objectives, hence the name 'signpost'. Ungemach et al. (2018) state that signposting effects are distinct from priming and framing effects, in the sense that they do not rely on valence shifts, but rather on a seemingly neutral conversion of different units. Furthermore, the effects of the study could not be explained by a knowledge gap.

As we can present the saving metrics in our energy recommender system with various units, e.g., Euro, CO2, or kWh savings, we are interested in the effect of these different signposts for users with different values. Following the study by Ungemach et al. (2018), we think that CO2 emission metrics might resonate more strongly with users who have stronger environmental concern. Similarly, monetary savings might resonate more with users who attach more importance to money, although this has not been studied before. We thus expect that a CO2 signpost will motivate users with a high level of environmental concern to save more energy. We also expect that the monetary (Euro) signpost will increase savings for users who place more importance on money, and that the kWh signpost will work equally well for everyone, as this has no direct apparent link to personal values. We also expect various effects on user experience, which we will elaborate on in the next sections.

- **[H2]:** *Signposts aligned with user values lead to higher savings and better user experience outcomes, than signposts that are not in line with user values.*
- **[H2a]:** *CO2 signposting, compared to kWh signposting, will result in higher savings and a better user experience*[1] *for increasing strength of people's environmental concern.*
- **[H2b]:** *Monetary signposting, compared to kWh signposting, will result in higher savings and a better user experience for increasing strength of people's financial values.*

We do note that a similar study involving signposting and goal setting has previously been conducted by Brandsma and Blasch (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

[1] 1 User experience consists of choice difficulty, system satisfaction and goal support (Subjective System Aspects), and environmental self-efficacy and choice satisfaction (Experience outcomes).

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, the study by Brandsma and Blasch (2019) only considered a single energy-saving measure: turning off stand-by appliances on a daily basis. Furthermore, participants were only asked to imagine setting themselves a goal and then say if they were willing to perform this action. Given that our current study differs substantially (with over 130 different saving measures, and persuasion towards real behaviour change), we hypothesised that these system manipulations might lead to different outcomes in our system.

User evaluation

Platforms that use recommender systems are usually evaluated on behavioural outcomes (Jannach et al., 2010), but several platforms also evaluate perception and experience aspects (Knijnenburg et al., 2014). Beyond achieving energy savings, users of an energy conservation platform should also be willing to re-use it (Geelen et al., 2019; Starke et al., 2017). Previous recommender studies show that the effects of system manipulations on user behaviour (i.e., amount of energy savings chosen) and user experiences (e.g., satisfaction) are mediated by user perceptions (Knijnenburg et al., 2014). For example, in a different energy-saving recommender study, the interaction effect between preference elicitation method complexity (i.e., the method of asking what a user likes and wants) and domain knowledge of users affected how satisfied users were and, in turn, how many energy savings they selected (Knijnenburg et al., 2014).

Such mediated relations are described by the user-centric evaluation framework by Knijnenburg et al., (2012, 2015). They show that system outcomes in recommender systems (e.g., choice satisfaction, kWh savings), are often explained by changes in subjective perceptions of the system itself ('subjective system aspects', such as choice difficulty and system satisfaction). In a study on the use of social norms in an energy recommender system (Starke et al., 2021), all effects of various social norms on number of chosen savings and choice satisfaction, were mediated by changes in perceived feasibility (a subjective system aspect). In a study on effective user interface designs, the effects of interface variations on choice satisfaction, were mediated by changes in perceived effort and perceived support (Starke et al., 2017). Although in this study, there were also some direct effects from interface variations on measure choices, which were not explained by changes in subjective system aspects.

In the current study, we want to evaluate the effects of goal setting and signposting on system outcomes, and if these effects exist, evaluate whether they can be explained by changes in how users perceive the system. We will explain this in detail in the conceptual framework that we will discuss in the next section.

Conceptual framework

Following the user-centric evaluation framework (Knijnenburg & Willemsen, 2015), we are interested in several system outcomes, as well as several subjective system aspects. For system outcome variables, these are the ways in which people evaluate their experiences of using the system (i.e., user experience), including the extent to which people are satisfied with their chosen measures (choice satisfaction), and the extent to which they now feel confident in saving energy (energy-saving self-efficacy). These also include more behavioural outcomes, such as the total selected savings (kWh savings, as an interaction outcome).

Following the framework's setup (Knijnenburg & Willemsen, 2015), we formulate a number of expectations. See Fig. 2 for an overview. We expect that the effect of our objective system aspects (goal setting and signposting) on the experience (self-efficacy and choice satisfaction) and behavioural outcome variables (selected kWh savings), will be mediated by subjective perceptions of the system. These include the difficulty users experience in choosing between measures (choice difficulty), the extent to which they feel the system supports their saving goals (goal support), the extent to which users are satisfied with the system (system satisfaction), and lastly, the extent to which users believe that the presented measures are feasible for them (perceived feasibility).

For example, presenting information in a way that is relevant to the user (i.e., signposting in line with user values), might aid the decision-making process, thus reducing choice difficulty and possibly improving goal support. We expect that such subjective perceptions of the system could lead to higher savings

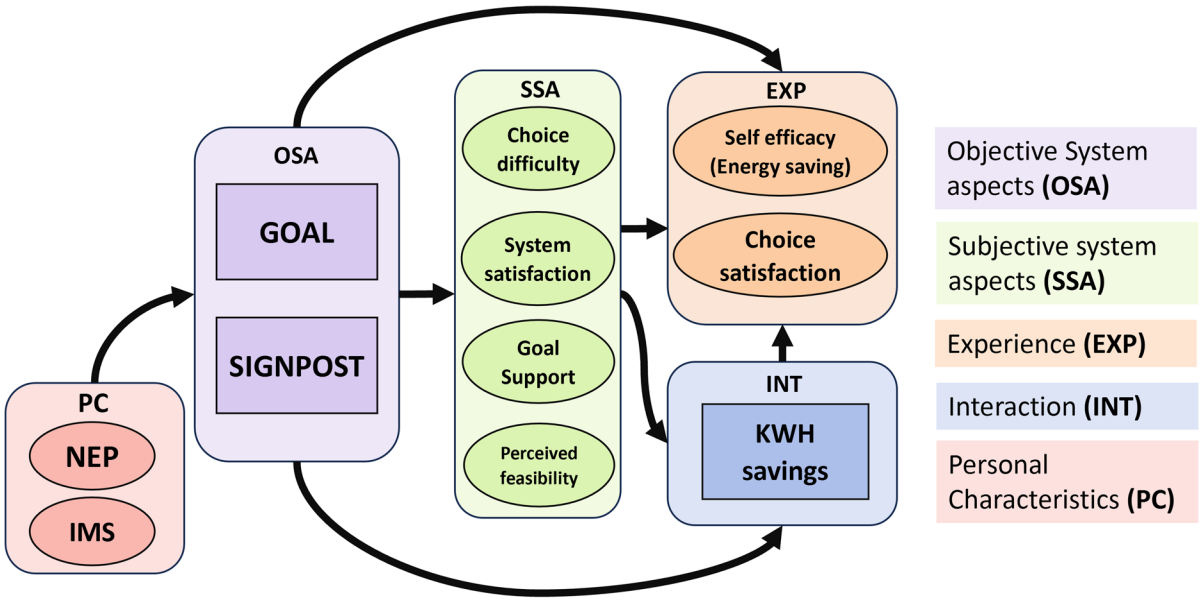


Fig. 2 Conceptual model, based on the evaluation framework of Knijnenburg and Willemsen (2015). We expect that the effects of objective system aspects on interaction and experience outcomes, will be mediated by changes in subjective system aspects, *NEP* = New Environmental Paradigm (environmental concern), *IMS* = Importance of money score

and improved user experiences. A similar trend was also seen in Starke et al., 2021. We furthermore think that goal setting could improve user experience by providing a clear behavioural target. Goals that are realistic and achievable for a user, might lead to higher system satisfaction, and in turn (following the framework), might lead to higher savings.

Lastly, increased savings might lead to increased choice satisfaction, similarly to the findings of Knijnenburg et al., 2014. Figure 2 depicts these hypotheses, in a model that broadly follows the user-centric recommender evaluation framework.

Figure 2 also presents additional paths, which follow the general route of objective system aspects to system outcomes via subjective system aspects, but of which the precise relationships are not based on earlier studies or theory (e.g., because we introduced our own measure of 'goal support', or our system differs from earlier ones in terms of system manipulations). Following the Framework for Evaluating Recommender Systems (FEVR) by Zangerle and Bauer (2022), and specifically their distinction between confirmatory and exploratory evaluations of recommender systems, we note that our current study, and especially the SEM model part of this study, has more of an exploratory character rather than a confirmatory one. The possible relationships between objective system aspects and eventual system outcomes might be explained through multiple possible pathways, however, not all of these pathways can be theoretically supported due to a lack of prior work.

We summarize this conceptual framework and our approach with the following (exploratory) hypothesis:

- **[H3]** The effects of objective system manipulations on interaction and experience outcomes are mediated by changes in subjective system aspects.

Methods

Our study consisted of an initial study where energy-saving measures could be chosen, and a four-week follow-up study, in which participants indicated which of their chosen measures they had performed (or started with). In an earlier study by Starke et al (2017), it was found that users who were more satisfied with their recommendations, also performed more items after four weeks. Especially for the goal-setting condition, we wanted to

check if a possible motivation to choose more measures, also translated into a motivation to perform more measures.

Participants

In total, 212 people participated in the initial study (202 after outlier removal), consisting of 86 males, 112 females, and 4 other/undefined, with a mean age of 30.5 and a median age of 25, with 75% of participants being younger than 32. 82% obtained at least some college education, while only 23% were homeowners. Out of all participants, 22% lived in a semi-detached 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 E or below.

We recruited participants at least 18 years old, living in the Netherlands, as the saving measures were tuned to the Dutch context. We excluded those below 18 as they might have particularly little say in household energy usage. 117 participants were recruited via our university departments' participant database. This database consists of (former) students and older (non-student) adults. The mean age of this group was 31.3 (95% $CI =$ 28.1. $SD =$ 34.5). 94 participants were recruited via the online Prolific database, which attracted a relatively young demographic to the study, with a mean age of 29.9 (95% $CI =$ 28.05, $SD =$ 31.84). We had one external participant who joined because someone sent them an invite in the final screen. Although we aimed our recruitment to contain mostly adult participants that would be able to engage in energy saving measures given their living situation, the sampling still resulted in a somewhat young sample, though still substantial older than the typical student sample and for most in living situations for which our energy savings were relevant.

The follow-up study was completed by 170 returning participants (160 after outlier removal), with a comparable demographic distribution. Studies were run between May and June 2023. Participants were paid ~€3,- for completing the initial study, and ~€4,- total if they also completed the follow-up, exact amount depending on participant database.

Study design

Our experiment was subject to a 2 × 3 between-subject design (Table 1). First, participants were either asked to set a goal or not. Second, there were three signpost conditions, presenting attributes and goals in terms of either kWh savings, monetary (€) savings, or annual kg CO2 reductions.

Procedure

The study used a newly developed online recommender system, based on earlier work (Starke et al., 2021), backed by a database of 135 energy-saving measures, obtained from Bams (2018). These measures ranged from 'easy' ones, e.g., *'Cook with pots & pans the same size as the heating element'*, to more difficult ones, e.g., '*Install a centralized temperature system with zone controls & thermostats*', to *'install a mini windmill'* as the most difficult measure. To be able to provide ability-tailored advice, users indicated for 19 measures (randomly selected from the entire range of measures) whether they performed them or not, referred to as '(number of) current items' in Table 3. This allowed us to calculate the ability of the user (which ranged between -2.7 and 3.6 logits for the entire sample). Measures that did not apply to a user's housing situation were not used in the ability (i.e., attitude) calculation.

Afterwards, participants in the goal condition could select an accessible (600 kWh), moderate (1200 kWh) or challenging (1800 kWh) saving goal, based on the mean savings (1200 kWh) and median savings (660 kWh, for those who selected at least some amount of savings in the study by Bams (2018)). This goal was displayed with the signpost a participant got assigned to (kWh/Euro/CO2). For the monetary goal, we used the weighted average of Dutch kWh and/m^3 prices (€0.30/kWh), resulting in saving goals of 180, 360 and 540 Euros. For the CO2 goals, we used an average of 0.25 kg CO2/kWh, resulting in saving goals of 150, 300 and 450 kg CO2. All other

Table 1 Study design

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

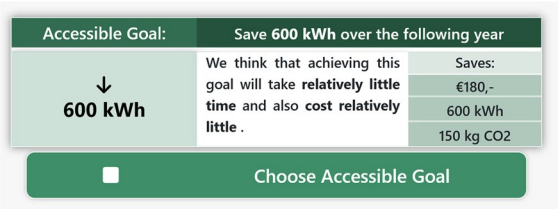


Fig. 3 Example of a possible energy-saving goal. On the left we see the signpost that the participant was randomly assigned to (in this case kWh). All other signposts were also included on the right. Next to this accessible goal, we had a moderate (1200 kWh) and a challenging (1800 kWh) goal

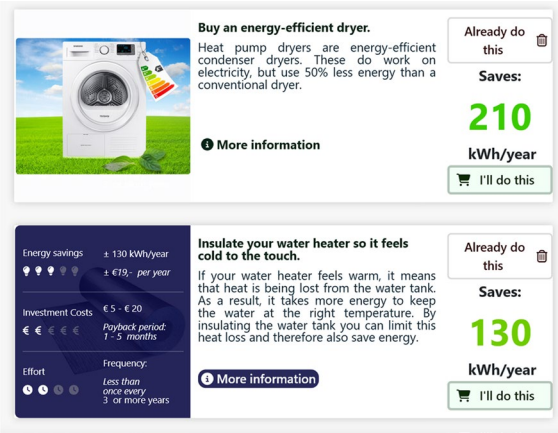


Fig. 4 Example items from the recommender list. The bottom measure depicts the information container, which is shown when clicking the 'more information' button or when hovering over the image

signposts were also shown as additional information. An example of a goal is given in Fig. 3 (600 kWh).

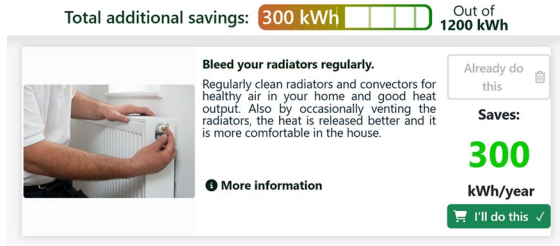
Next, participants were given instructions on how to use the interface with screenshots (Refer to Oonk, 2023, Fig. 6). They were asked to pick items they intended to perform in the future, to remove items they already performed (after which new ones would appear at the bottom of the list) and were shown how to obtain more information on measures (by hovering over the images or clicking the 'more info' buttons). An example of the interface is given in Fig. 4.

Afterwards, all participants were presented 20 energy-saving recommendations closest to their ability level (sorted by absolute distance) to choose from. (We refer to the total number of chosen items as '# chosen items in Table 3, and the total amount of savings (e.g., in total kWh or total monetary value), as chosen savings throughout this paper.) The measures were accompanied by highlighted saving metrics as seen on the right in Fig. 5, which were translated as either a kWh, monetary, or CO2 metric depending on the signpost condition the participant was in.

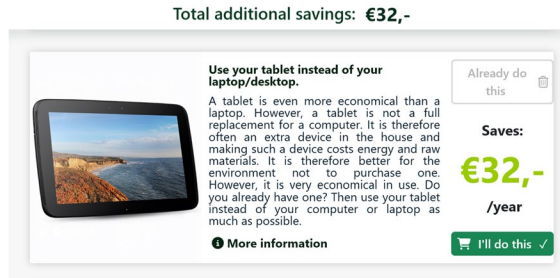
For exact conversion formulas between these, refer to Oonk (2023). The total selected savings were shown at the top of the screen, as a proportion of the saving goal if applicable (as depicted in Fig. 5a and Fig. 5b). The recommender interface can be tested at: www.besparingshulp.nl/demo.

Measures

Following the choice task, participants were surveyed about their perceptions and experiences with the system, with the questions shown in Table 2. We used items from Willemsen et al. (2016) to measure choice difficulty, items from Starke et al. (2017) to measure perceived feasibility, items from Willemsen et al. (2016) to measure choice satisfaction, and items from Knijnenburg et al., (2012, 2014) and Starke et al. (2017) to measure system satisfaction. Goal support



a. Goal condition



b. No-goal condition

Fig. 5 (Partial) screen captures of our recommender system interface. Depicted in **a** is our Goal (kWh) condition, and in **b** the no-goal (euro) condition. The progress bar in the goal condition has a logarithmic scale

Table 2 Questionnaire items and factor loadings (λ) for both the initial (1) and follow-up (2) samples. Alpha denotes Cronbach's Alpha, AVE the Average Variance Explained. Choice difficulty and choice satisfaction questions were obtained from Willemsen et al. (2016), and goal support questions were largely original questions to measure the extent to which participants felt the system helped them save energy

Construct	*Name*	*Proposition*	*Study*		*follow-up*	
			λ	R^2	λ	R^2
Choice difficulty	Cdif1	It was easy to choose between energy saving measures	−0.73	0.56	−0.83	0.71
Alpha:0.66(1)/0.67(2)	Cdif3	The task of choosing energy saving measures was overwhelming	0.53	0.30	0.50	0.28
AVE: 0.45(1)/0.47(2)	Cdif4	Comparing the energy saving measures took a lot of effort	0.73	0.50	0.63	0.43
Choice satisfaction	Chsat1	I am satisfied with the measures I chose	0.49	0.66	0.51	0.78
Alpha: 0.73(1)/0.76(2)	Chsat2	I think I would enjoy performing the chosen energy saving measures	0.41	0.47	0.39	0.46
AVE: 0.57(1)/0.65(2)	Chsat3	I would recommend the chosen measures to others	0.47	0.59	0.49	0.72
Goal support	Syssat2	The Saving Aid (SA) is helpful to find appropriate measures	0.74	0.62	0.70	0.51
Alpha: 0.86(1)/0.82(2)	Gsup4	The SA gives me more insight into the energy consumption of devices and systems in my home	0.77	0.68	0.78	0.63
	Gsup5	The SA makes me more energy-conscious	0.77	0.69	0.78	0.63
AVE: 0.68(1)/0.62(2)	Gsup6	The SA makes me more aware of my options for saving energy	0.78	0.71	0.83	0.72

questions were largely original items, measuring how well the system supports the pursuit of energy saving.

Additionally, we used the revised New Environmental Paradigm (NEP) (α: 0.81 for both the initial study and follow-up study) by Dunlap et al. (2000) to measure environmental concern, the Money Importance Scale (IMS) by Franzen and Mader (2022) (α: 0.77 for both samples) and the environmental self-efficacy scale by Lee and Tanusia (2016) (α: 0.87 & 0.85, AVE: 0.60 & 0.57 for the initial and follow-up study respectively). For the full overview of questions (before factor analyses), refer to Oonk (2023). To verify the proposed constructs of our theoretical model, we performed an exploratory factor analysis (EFA) and then CFA and SEM using the mPlus software package (Muthén & Muthén, 2023). Items with low factor loadings (below 0.4) or high cross-loadings in the EFA (main loading of a factor should be twice as high as the loadings on another factors) or in CFA/SEM, items with several high (> 20) Mod indices with other items (Several goal support, choice satisfaction, and system satisfaction questions) were removed from the subsequent CFA and SEM analyses. In our SEM analysis, Perceived feasibility had a too high correlation with choice difficulty (correlation above 1, indicating in MPlus that these factors cannot be discriminated in the SEM model). The predictors of perceived feasibility were weaker than those of choice difficulty. Additionally, our system manipulations were more focused on helping users choose (by emphasizing reasonable saving targets and providing personally meaningful information through signposts), rather than influence perceived feasibility; After all, users chose their own measures (and their own goal, when applicable), and thus partially determined themselves how feasible their final set of measures would be. Therefore, we included choice difficulty in our final SEM model rather than perceived feasibility. 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. Participants could save their selected measures and received an email with an overview one week after the study. After four weeks, participants were asked to join the follow-up study to indicate which measures they ended up performing.

We compared the differences in savings between goal conditions with rank-sum tests, the interaction effect between values and signposts with several (robust) multiple regressions, and the effect of subjective system aspects with a structural equation model (SEM).

Results

We found that participants evaluated the system very positively. 89% agreed at least to some extent with the proposition that "the saving aid was helpful to find appropriate measures". Participants chose on average 7.5 new measures in the initial study ($M = 3057$ kWh, SD $= 4406$), of which they reported to complete on average 2.1 measures four weeks later ($M = 315$ kWh, $SD = 538$). To see if higher chosen savings indeed lead to saving more energy in the self-reported measures taken four weeks after the study, we compared chosen and actual savings from the initial study and the follow-up study, respectively. This correlation can be seen in Fig. 6. According to a pairwise correlation, after a log transform, this relationship was significant with a coefficient of $\beta = 0.36$ and $p < 0.0001$. This means that people who choose a higher amount of savings, indeed end up saving more energy.

On average, participants removed 10.3 measures from the recommendations, stating that they already performed them. We found no correlation between NEP score and ability level, suggesting that those with stronger environmental concern did not report to already perform more actions. We also did not find a correlation between the chosen goal amount and either NEP or IMS scores (Refer to Table 3). The NEP scale and IMS scale both had 5-point Likert scale answer options, resulting in a possible range for computed scores from −2 to 2. For the NEP scale, we found an average score of 0.84 with a standard deviation of 0.55, a minimum of −0.5 and a maximum of 2. For the IMS score (in the initial study), the average score was 0.24, with a standard deviation of 0.72, a minimum of −1.7 and a maximum of 2. Refer to Fig. 7. We did observe that NEP score and IMS score were slightly negatively correlated (−0.25). As we invited participants in batches, some completed the initial study and follow-up study within a shorter timespan than others, ranging from around 19 days between the two studies to more than 40 days. However, we saw no significant correlation between this duration and the achieved savings at the follow-up study (Coefficient = 0.09, $p = 0.28$).

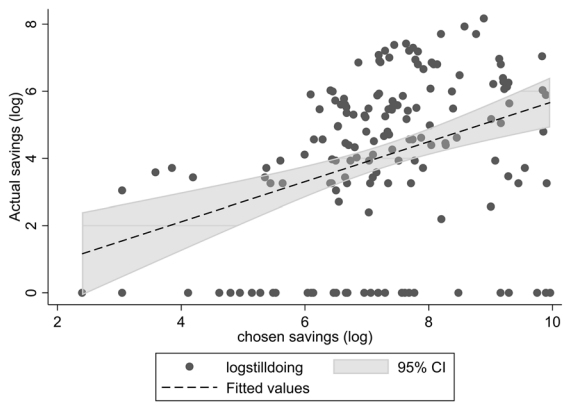


Fig. 6 Correlation between chosen savings and actual savings after four weeks (not concerning those with chosen savings of 0 kWh)

Goal conditions

We first compared the savings between the goal and no-goal conditions. We hypothesized [H1] that users in the goal condition would obtain more savings than users in the no-goal condition. The energy savings for the initial (main) study and the follow-up survey across these conditions are depicted in Fig. 8. On average, surprisingly, participants chose less energy savings in the goal condition ($M = 2166$ kWh; $SD = 2624$) than in the no-goal condition ($M = 3880$ kWh; $SD = 5452$). A rank-sum test indicated that this difference was, however, not significant. Four weeks later, in the follow-up survey, the difference was reversed, with goal-condition savings being higher on average ($M = 340$ kWh; $SD = 565$) compared to the no-goal condition ($M = 294$ kWh; $SD = 516$). However, once again, this difference was not significant ($p = 0.46$).

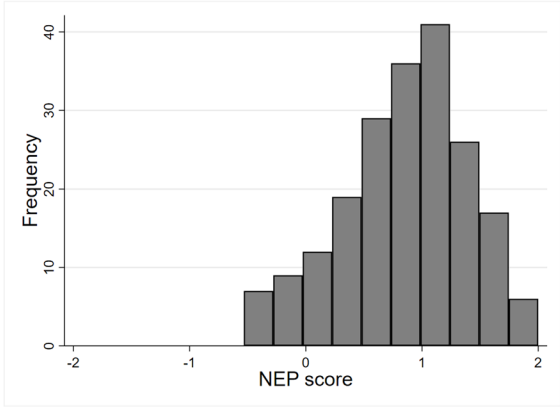
Goal Amounts

We furthermore compared the savings between the three goal amounts, as graphed below in Fig. 9. 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.

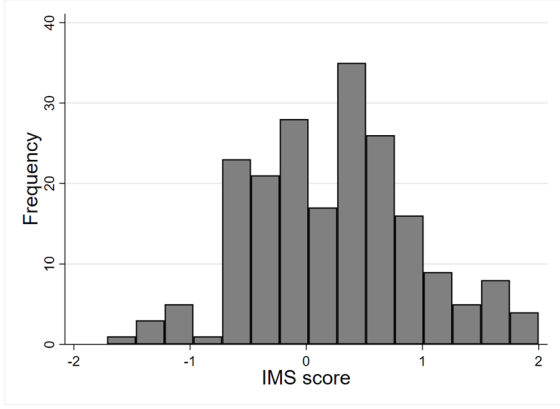
Table 3 Pairwise correlations for person characteristics in initial study. (N = 202)

Variables	(1) NEP	(2) IMS	(3) Male	(4) Age	(5) kWh	(6) home	(7)En. lab	(8) Cur	(9) new	(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) chosen 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) #Chosen 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 $p < 0.05$ **Was significant with a coefficient of 0.30 when considering only the goal condition



a. NEP scores distribution



b. IMS scores distribution

Fig. 7 Distributions of NEP (**a**) and IMS (**b**) scores for the initial sample without outliers

Signpost conditions

Figure 10 compares the chosen savings per signpost condition. We hypothesized [H2] that signposts in line with user values, would result in higher savings than signposts not in line with user values. First, we compare the savings in each condition, irrespective of user values. We found that people chose significant less savings in the CO2 condition in the initial study (M = 1782 kWh; SD = 2554), compared to the kWh (M = 3652 kWh, SD = 5049) and Euro (M = 3620 kWh, SD = 4843) conditions ($\chi^2 = 10.11$, $p < 0.001$). After four weeks, participants saved on average 347 kWh (SD = 507), 334 kWh (SD = 573) 268 kWh (SD = 538) in the CO2, kWh, and Euro conditions, respectively. An Analysis of Variance (ANOVA) showed no significant difference between these three signpost conditions in the follow-up study, nor when a quadratic transformation was performed.

Signposts and values

To test for the signposting effects [H2], we inspected how a user's values (NEP) affected behavioural and user experience outcomes, using moderated regression. See Fig. 11. We found a positive interaction effect between the NEP score and a CO2 signpost, where a higher NEP score led to higher choice satisfaction ($\beta = 0.30$, $p < 0.01$) and higher self-efficacy ($\beta = 0.35$, $p < 0.05$), compared to the downward-sloped

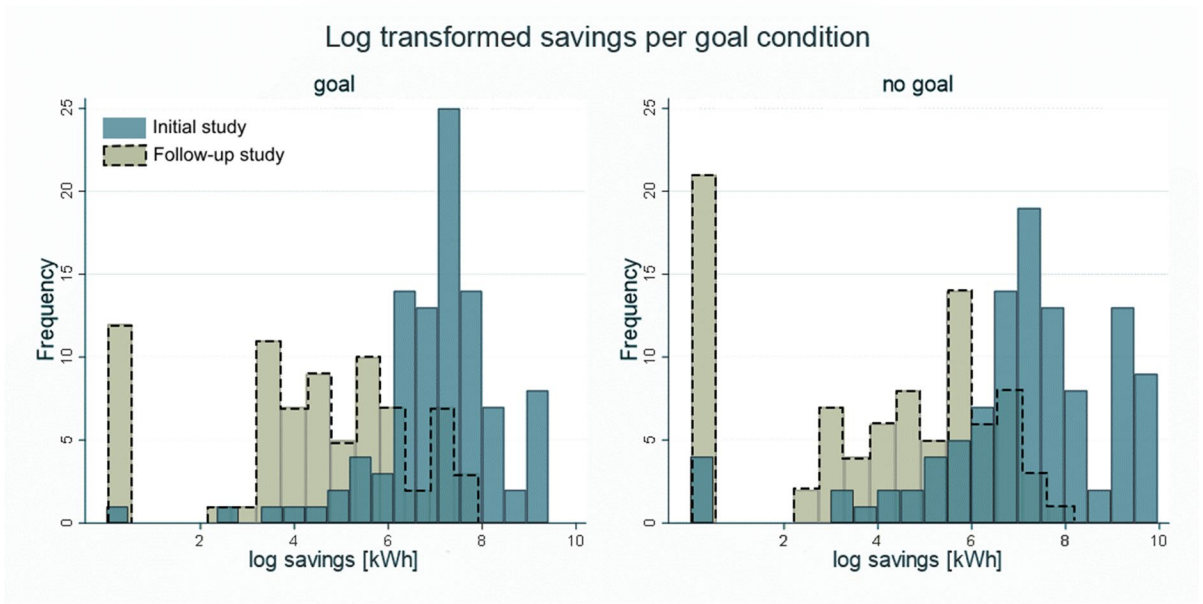


Fig. 8 Chosen (initial study) and performed (follow-up study) Savings in the goal and no goal condition

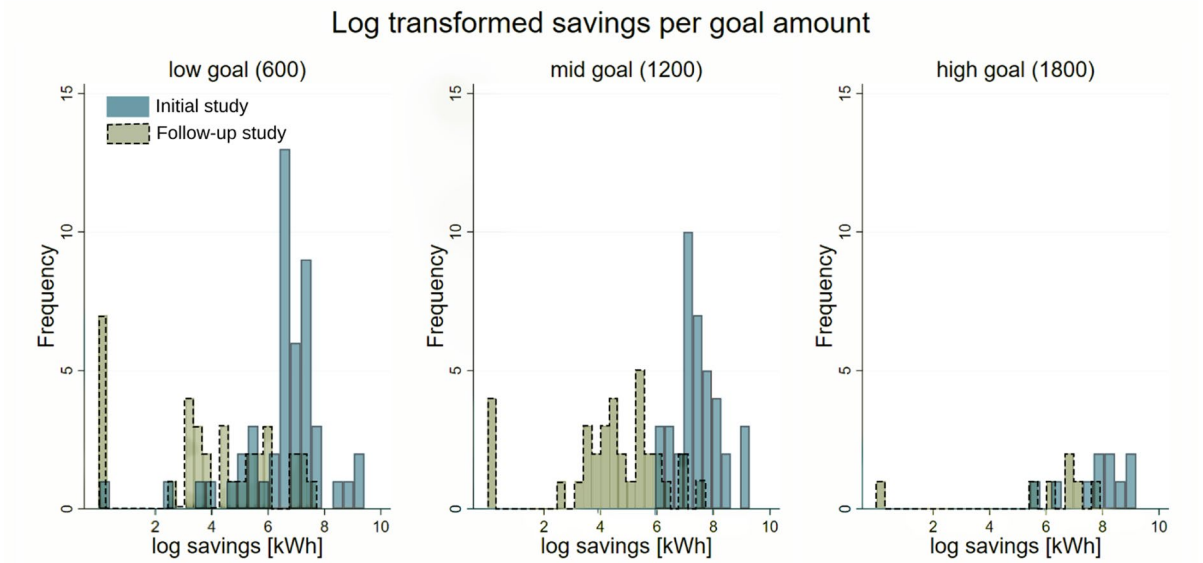
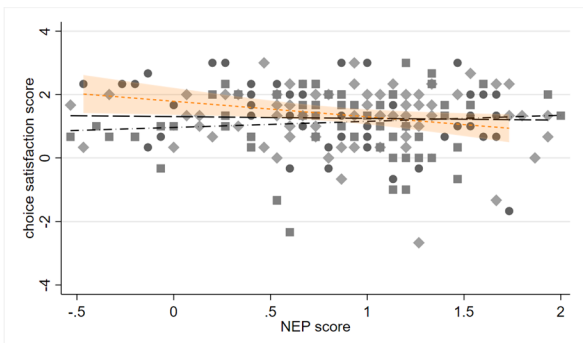


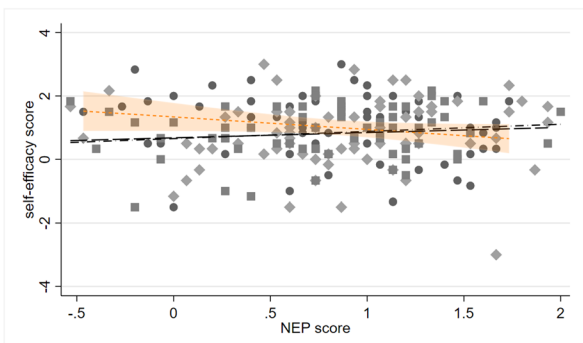
Fig. 9 Chosen (initial study) and performed (follow-up study) savings per chosen goal amount

effect of the kWh signpost baseline. The latter is depicted by the orange lines in Fig. 11a and b. However, this was mostly due to a downward slope of the kWh signpost, as depicted by the blue lines in Fig. 11a and b. For choice difficulty, the interaction effect of NEP and CO2 was negative ($\beta = -0.55$, $p < 0.01$, Fig. 11c). Again, the kWh baseline differed between low and high NEP scores more strongly than the CO2 signpost did. Additionally, the main effects of the CO2 signpost on choice satisfaction ($\beta = -0.59$, $p < 0.01$)

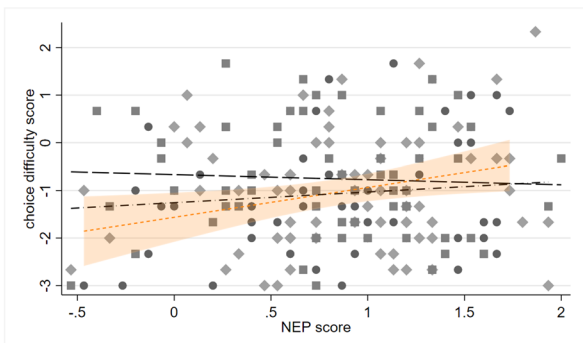
Fig. 10 Savings per signpost condition, for both the initial (main) study and the follow-up study



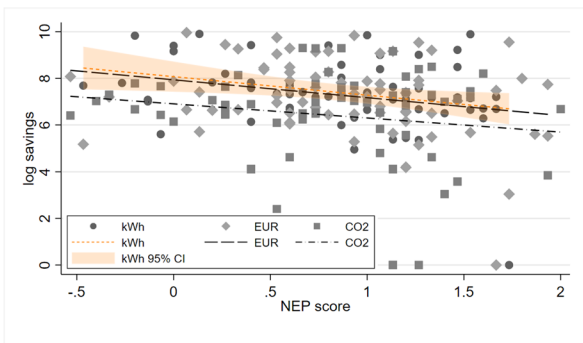
a. NEP score and choice satisfaction score.



b. NEP score and self-efficacy score



c. Nep score and choice difficulty score



d. NEP score and log chosen savings

Fig. 11 Effects of the NEP score against the: choice satisfaction score (**a**), self-efficacy score (**b**), choice difficulty score (**c**), and savings (**d**) for different signposts. The blue shaded area depicts the 95% confidence interval of the kWh signpost

Table 4 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]
EU	−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			

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

and self-efficacy ($\beta = -0.45$, $p < 0.05$) were negative (see the regression outputs in Table 4). Therefore, the beneficial effects of a CO2 signpost only applied to a minority with very high NEP scores, while the kWh signpost seems beneficial for a larger group of people lower on the NEP scale. We did not find such interaction effects on actual energy savings in the initial and follow-up studies (Fig. 11d, depicts the savings in the initial study). We neither found any interaction effects between signposts and the Importance of Money (IMS) score on any of these outcome variables.

Subjective system aspects and SEM model

Finally, we hypothesized that the effects of system manipulations (goal setting and signposting) on system outcomes, would be mediated by changes in subjective system aspects [H3]. To examine these mediated effects between changes in the interface, user perceptions and choices, and system evaluation, we constructed a structural equation model (SEM) in Mplus (Muthén & Muthén, 2023). The SEM model had a good fit: $\chi 2(202) = 255.00$, $p < 0.01$, $RMSE = 0.040$, 90%-CI $= [0.025, 0.052]$, $CFI = 0.981$, $TLI = 0.979$. The fit for the follow-up SEM model was also good, and their results are merged in Fig. 11, depicting only significant paths. The system satisfaction variables were only measured in the initial study, and their effects remained very similar in the follow-up. The paths from Euro signpost were no longer significant in the follow-up study. For separate SEM graphs of the initial and follow-up study, refer to Oonk (2023; p. 68, Fig. 18) and Oonk (2023; p. 88, Fig. 28).

We found several correlations between signposts and latent variables, some of which were mediated by pro-environmental values (NEP score). We found no direct effects from signposts and interactions on outcome variables (Savings, choice satisfaction, and self-efficacy), nor did we find an effect of the Importance of Money Score (IMS score) on savings or any of the other factors. Discriminant validity of choice satisfaction (AVE 0.57) is not maintained, because choice satisfaction was better explained by choice difficulty ($\beta = -0.81$, $p < 0.001$) than by its own questions. Choice difficulty, as per the user-centric evaluation framework by Knijnenburg and Willemsen (2015), is considered to be a subjective system aspect, and the questions we used measure user perceptions, whereas choice satisfaction is usually referred to as an experience outcome and the questions we used are more

evaluative. Moreover, they show different (and logical) relations to the other related concepts (self efficacy and goal support). 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 and other concepts in our model.

From the model in Fig. 12, we observe that all possible effects from signposts and values on these outcome variables were mediated by choice difficulty, and all effects on savings were mediated by changes in goal support. There were no direct effects from any of the objective system aspects nor the personal characteristics (NEP), on outcome variables. We expected that all effects on outcome variables (savings, choice satisfaction and energy-saving self-efficacy), would be mediated by changes in subjective system aspects. While the main effect of a CO2 signposts (compared to a kWh signpost) *increased* choice difficulty ($\beta = 1.42$, $p < 0.001$), CO2 signposts moderated by higher NEP scores *decreased* choice difficulty ($\beta = -0.55$, $p < 0.01$). Effects on savings were indeed furthermore mediated by goal support; the reduction of choice difficulty, led to increased experienced goal support ($\beta = -0.38$, $p < 0.001$). This increased goal support, then led to higher chosen savings in the initial study ($\beta = 0.49$, $p < 0.001$), and higher actual savings in the follow-up study ($\beta = 0.71$, $p < 0.001$). The total indirect effect of the interaction CO2 x NEP score on (initial) savings via this route was $\beta = 0.10$ ($p < 0.01$), meaning that increasing NEP score led to slightly higher savings in the CO2 condition as compared to the kWh condition, and this was caused by a decrease in choice difficulty and increase in goal support. This is in line with hypothesis 2a, but something we did not observe in the regression in Sect. 3.3, possibly because more variables (IMS scores and goal conditions) were included there. However, if we would correct for this double testing, this interaction effect would still be significant ($p < 0.025$).

Next to a mediation of subjective system aspects on savings, we expected similar mediation effects on experience outcomes (energy-saving self-efficacy and choice satisfaction). Indeed, a decrease in choice difficulty led an increase in self-efficacy ($\beta = -0.38$, $p < 0.01$). The total indirect effect from the interaction NEP x CO2 on self-efficacy, including the 3 other paths via goal support, savings and choice satisfaction, was $\beta = -0.97$ ($p < 0.01$). Increased goal support also led to higher choice satisfaction ($\beta = 0.57$, $p < 0.001$). Lastly, increased selected savings led to an increase

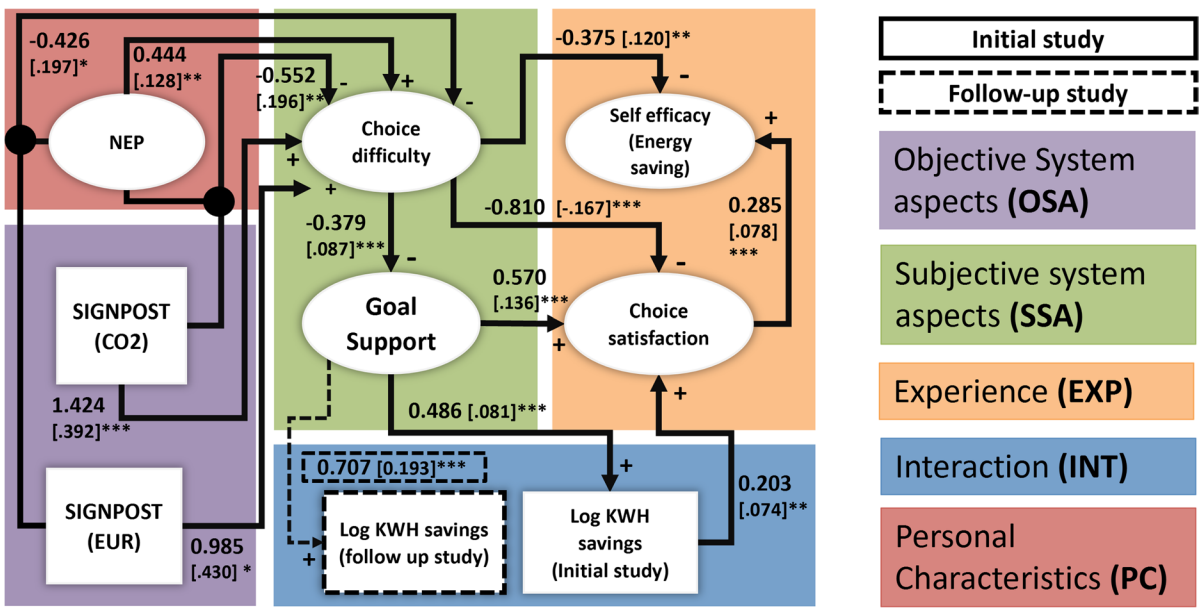


Fig. 12 Structural Equation Model (SEM). Numbers on the arrows represent the β-coefficients; standard errors are between brackets. Effects between subjective constructs are standardized and resemble correlations. ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$. Dots depict interaction effects. Dotted line depicts a path of the follow-up study

in choice satisfaction ($\beta = 0.20$, $p < 0.01$), which in turn led to a higher energy-saving self-efficacy ($\beta = 0.29$, $p < 0.001$). This all shows that indeed, these experience outcome variables were also mediated by changes in subjective system evaluations, similarly as seen with the savings, with the most important mediators being choice difficulty and goal support.

As choice satisfaction and self-efficacy are outcome variables that we measured only in the initial study, the pathways from actual savings in the follow-up study, to these outcome variables, are not depicted in the figure. They can be found in Oonk (2023).

Discussion

We have evaluated the effectiveness of goal setting and signposting on energy-saving choices and user experience in the context of a tailored, Rasch-based recommender system. Our Rasch scale of energy-saving measures, rooted in Campbell's Paradigm, helps to effectively assess the ability of users, allowing our recommender system to present tailored advice. Our study has examined short-term user preferences for these tailored household energy-saving measures, examining user choices, perceptions, and evaluations, as well as users' self-reported saving behaviour four weeks later. We have found that the system is rather effective for household energy conservation, with users reporting an average 316 kWh predicted yearly savings (both gas and electricity combined) per person after four weeks. However, while users report to appreciate the system, we have found mixed results regarding our goal-setting and signposting manipulations. We did observe that all effects from signposting on outcomes were explained by changes in perceived choice difficulty and goal support, in line with some of our expectations and earlier work of Knijnenburg et al. (2014)

Conceptual framework

We have observed various indirect effects of system characteristics on subjective perceptions of the system and have observed that these subjective perceptions influenced experience and behaviour outcomes. We have furthermore observed that decreased choice difficulty led to increased goal support (both subjective measures), and that increased goal support led to higher savings and higher choice satisfaction. 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).

Goal conditions

We have not observed significant differences in kWh savings or user experiences between the goal and no-goal conditions, in contrast to previous non-recommender studies in an energy-saving context (Abrahamse et al., 2007; Harding & Hsiaw, 2014). We argue that the selection of energy-saving measures by users could have been experienced as a goal-setting task in itself (i.e., the total selected measures constitute a saving goal), diminishing the effect of overarching goal setting. Furthermore, the personalized Rasch system already aids users in finding ability-matched saving measures, which could minimize the effects of goal setting in terms of decision support. Additionally, goal setting in an energy-saving context is often more effective when paired with feedback (Abrahamse et al., 2005), which was lacking in our session-based system. Goal setting within a system that can be used continuously over a longer time, might be more effective. Although participants received a link to their chosen measures approximately one week after participation, the goal was no longer visible after the initial study. This might have made the goal setting less salient and less impactful on the behaviour of participants.

Lastly, participants selected on average 3057 kWh in savings, as compared to 1200 kWh in a previous study (Bams, 2018), surpassing the levels of the goals we suggested. This might have been due to a misinterpretation of the task for some participants: Several participants asked why there was not a 'I do not want to do this' button, and 33 participants clicked on either 'I'm already doing this' or 'I will do this' for all items in the list. They most likely thought that they had to click on either of those buttons for every recommendation in the list, while this was not required (they were allowed to 'ignore' recommendations entirely, though we should have made this clearer in the instructions). This happened 13 times in the goal condition, and 20 times in the no-goal condition. This misinterpretation might have had a stronger effect in

the no-goal condition, due to the lack of a realistic saving target. For the signpost conditions, this happened 12 times in the Euro condition, 14 times in the kWh condition, and 7 times in the CO2 condition. Furthermore, we had 14 participants (across conditions), click on all but 1 or 2 items in the list, who might have had the same misunderstanding. Excluding these 33 or 47 participants would bring the average chosen kWh savings down to 2257 kWh or 2073 kWh respectively.

Signposting

Furthermore, we have compared the effects of three different signposts (kWh, CO2, and Euro) on user energy-saving behaviour and user experience. We have found that the CO2 signpost has led to lower chosen savings in the initial study, irrespective of value orientation. However, there was no significant difference in achieved savings after four weeks. It could be that the kWh and Euro signposts motivated people to choose more measures, but this did not lead to a higher propensity to perform these measures. We have observed that users facing a CO2 signpost reported better user experience outcomes at the upper end of the NEP scale, compared to users facing a kWh signpost. This was, however, caused by the kWh signpost being less compatible with increasing NEP score, rather than the CO2 signpost working better with increasing NEP scores.

This contrasts with the findings of Ungemach et al. (2018), where a CO2 signpost led to more efficient car choices with increasing NEP scores. This car comparison task was merely a choice task and did not call for real-world action. Similarly to the NEP scale, such hypothetical measures might not always correlate with real-world behaviour. Previous studies have shown that a higher level of environmental concern (NEP), does not lead to an increase in energy savings (Urban & Ščasný, 2012). The energy recommender system we used, calls for an actual change in behaviour, and might therefore have yielded different results.

Savings

Although not the main aim of our current research, the eventual aim of the platform is to help people save energy. Participants achieved on average (316 kWh) of predicted yearly savings with the measures implemented after just four weeks. This might increase over time, as it does not include the items that a user says they have started implementing ($M = 349.8$ kWh; $SD = 1080.3$) or are still planning to do in the future ($M = 1314.8$; $SD = 2852.8$), only the ones they are already doing consistently or have fully implemented after four weeks.

Although we did not obtain data on the energy usage of our participants, we can try to put this 316 kWh into perspective; The average yearly energy usage of a Dutch household was 2500 kWh and 820 m^3 gas total in 2023 (Centraal Bureau voor de Statistiek (CBS), 2024). Converting gas savings to kWh savings (0.102 m^3 gas per kWh) in the same way we did in our system (Refer to Oonk, 2023), this would be 2500 kWh + 820/0.102 = 10,539 kWh equivalent energy usage on average. Although 316 kWh is then a relatively small percentage of the total amount of energy used (2.99%), it is already more than 10% of the electricity usage. We also note that out of the 10 most popular completed measures, 7 were electricity based, though the few gas-based measures generally had higher savings. Most measures that save substantial gas-usage like insulation, need more time to be implemented. We think that these predicted yearly kWh savings are still a worthwhile amount, especially for our somewhat younger sample. Additionally, it would be worth exploring how the system can support the not-yet-implemented but intended behaviours over the long-term.

Limitations and future research

Our study is subject to a few limitations. One notable demographic limitation is that of the sample's age. Our participants are relatively young, with a mean age of 30.5, of which 75% were below 32 years old. Few of our participants owned homes, meaning that some might not have been able to implement the often more efficient structural changes (as compared to behavioural measures). Therefore, participants might have been motivated to choose measures that they could not actually perform, and perhaps even to report completing them. We did not check whether non-homeowners picked items that would require structural changes, which would be interesting to look at for future research. Actual energy usage data might have been more informative than the self-reporting in

our system, though this would require a much longer-term study and might be experienced as impeding on privacy. Additionally, one of the appeals of a Rasch-based system, is precisely that one does not need to provide energy usage data to receive recommendations. Additional research that involves mostly homeowners or where certain items are filtered out for non-homeowners, might provide additional insights and could show clearer differences between conditions.

Furthermore, the list of measures had a large range of savings, from 0 to 8000 kWh. Although we did observe substantial differences in means between groups, these were not significant. This might, to a certain extent, also have been caused by large standard variations in our data. A larger sample size could indicate whether this larger variation is systematic. Making the saving aid publicly available to a broad audience, might make it easier to test certain system manipulations on a larger sample, and give further insight into possible effects. System manipulations would then be done by means of A/B testing, for example (presenting one set of users with one version, and another set of users with another). Participants furthermore had seemingly high levels of environmental concern, which might have influenced our results; 91% of participants agreed (to some extent) to the NEP scale proposition that *'Humans are severely abusing the environment'*. Thus, the observed range of NEP scores was perhaps too small to show large effects between groups. We observed an average NEP score of 0.84 with an SD of 0.55 (on a scale of −2 to 2, which would be 3.55 on a scale of 1–5). For comparison, a previous study in 2023 across Germany and Poland reported a mean of 0.695 (or 3.695 on a scale of 1–5), with a standard deviation of 0.542, with a somewhat older sample (Bohdanowicz et al., 2023). This is somewhat comparable to what we observed. Still, involving a wider range of users might lead to different results. The same goes for involving users from different countries: the results we found here are based on a Dutch sample, and the recommendations were also based on earlier research within the Dutch population. Such a system might show different results in different environmental and socio-economic conditions. For future research, it might also be valuable to look at user satisfaction after an extended period of recommender system use and measure implementation. Similarly, it would be interesting to compare self-efficacy before and after using the system, as self-efficacy variation likely exists in the sample already and might influence choices and behaviour up-front. In the current study, we assume that our system influences self-efficacy, but the level of self-efficacy prior to the study might also influence system usage and choice outcomes. Lastly, it might give additional insights to research a similar system that not only involves behaviours and investments within the living arrangements, but also those outside, e.g., green travel choices and purchasing behaviours.

In a broader technological context, recommender systems can be part of a smart home environment. Feedback on one's current energy use could be an important additional behavioural determinant of energy conservation (Fischer, 2008), particularly when it is instructive about which appliances to turn off (Fensel et al., 2014; Geelen et al., 2019).

Conclusions and implications

Our system helped users save a non-trivial amount of energy after just 4 weeks, and users were quite positive about the system in general. Given this, we see opportunities for the further usage and exploration of the system. This can be in the household energy saving domain, for other (sustainable) behaviours such as purchase behaviour, or energy usage in an industrial context. Although we think the system itself remains promising, the current study does not find support for the idea that goal setting would be effective to increase savings or improve user experience in a Rasch-based energy recommender system. However, we also did not find evidence for detrimental effects, which raises the question whether we would find different outcomes with an improved system design. Given the large body of previous research on goal setting, optimizing our goal levels, or providing users with direct feedback in a more interactive system, might lead to valuable additional insights. It should be considered, however, that no goal might be better than a goal not reached, in terms of system satisfaction, and that personalized recommendations might already provide sufficient support to the user to not require explicit goal setting.

We did find several effects of signposting on user experience. 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 improved energy-saving self-efficacy, reduced choice difficulty, and increased choice satisfaction. However, this was a rather small effect, and besides, the group of people for whom this would make a real difference (on the lower end of the NEP scale) would be rather small and possibly hard to reach. Therefore, we are unsure if signposting would be a beneficial addition to the current system.

We can state that, in general, participants evaluated the system positively, and that they achieved predicted yearly savings (316 kWh) after just four weeks, with a substantial number of measures that users still planned on doing in the future. We believe that these findings combined, give grounds for the further exploration, improvement, and usage of this system.

Acknowledgements We appreciate the efforts of all involved, including the study participants, those who shared feedback and insights at various stages in the process, and the reviewers and editors who helped us improve this manuscript.

Authors' contributions This paper is based on the MSc thesis of AO, the first author. MW and AS served as supervisors during this thesis project. All authors contributed to study conception, (recommender) design, and analysis plan. System implementation and data collection were done by AO, as well as data analysis with the help of MW. The first draft of the manuscript was written by AO, substantially revised by AS, and further adapted by MW. All authors read and approved the final manuscript.

Funding The study was done without specific grants or funding.

Data availability The datasets presented in this study can be found in the OSF online repository (for participants who opted in to public data sharing): https://osf.io/7qdgw/. Full datasets are available from the authors on reasonable request.

Declarations

Ethical approval and consent to participate, human ethics, and consent for publication. Participants signed an informed consent prior to participating, informing them of the goal of the study. The study was reviewed and approved by the ethical review board of the Human-Technology Interaction Group, Eindhoven University of Technology, The Netherlands.

Conflict of interest The authors declare that they don't have a conflict of interest relevant to the research, nor to the submission to this journal.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology, 25*(3), 273–291. https://doi.org/10.1016/j.jenvp.2005.08.002

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

Adaji, I., & Adisa, M. (2022). A review of the use of persuasive technologies to influence sustainable behaviour. In *Adjunct proceedings of the 30th ACM conference on user modeling, adaptation and personalization* (pp. 317–325). https://doi.org/10.1145/3511047.3537653

Alsalemi, A., Sardianos, C., Bensaali, F., Varlamis, I., Amira, A., & Dimitrakopoulos, G. (2019). The role of micro-moments: A survey of habitual behavior change and recommender systems for energy saving. *IEEE Systems Journal, 13*(3), 3376–3387. https://doi.org/10.1109/JSYST.2019.2899832

Attari, S. Z., DeKay, M. L., Davidson, C. I., & Bruine De Bruin, W. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences, 107*(37), 16054–16059. https://doi.org/10.1073/pnas.1001509107

Attari, S. Z., & Rajagopal, D. (2015). Enabling energy conservation through effective decision aids. *Journal of Sustainability Education, 8*, 1–15.

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 January 28, 2023, from https://pure.tue.nl/ws/portalfiles/portal/107655057/Master_s_Thesis_Luc_Bams_0804795.pdf

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

Bohdanowicz, Z., Łopaciuk-Gonczaryk, B., Gajda, P., & Rajewski, A. (2023). Support for nuclear power and proenvironmental attitudes: The cases of Germany and Poland. *Energy Policy, 177*, 113578–113578. https://doi.org/10.1016/j.enpol.2023.113578

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

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

Centraal Bureau voor de Statistiek (CBS). (2024, August 15). *Energieverbruik particuliere woningen; woningtype en regio's*. Centraal Bureau Voor de Statistiek. Retrieved March 14, 2025, from https://www.cbs.nl/nl-nl/cijfers/detail/81528NED

Eagly, A. H., & Telaak, K. (1972). Width of the latitude of acceptance as a determinant of attitude change. *Journal of Personality and Social Psychology, 23*(3), 388. https://doi.org/10.1037/h0033161

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

Fensel, A., Kumar, V., & Tomic, S. D. K. (2014). End-user interfaces for energy-efficient semantically enabled smart homes. *Energy Efficiency, 7*, 655–675. https://doi.org/10.1007/s12053-013-9246-2

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

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

Gardner, G. T., & Stern, P. C. (2008). The short list: The most effective actions US households can take to curb climate change. *Environment: science and policy for sustainable development, 50(5),* 12–25. https://doi.org/10.3200/ENVT.50.5.12-25

Geelen, D., Mugge, R., Silvester, S., & Bulters, A. (2019). The use of apps to promote energy saving: A study of smart meter–related feedback in the Netherlands. *Energy Efficiency, 12*(6), 1635–1660. https://doi.org/10.1007/s12053-019-09777-z

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

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). *Recommender systems: An introduction*. Cambridge University Press.

Kaiser, F. G., Byrka, K., & Hartig, T. (2010). Reviving Campbell's paradigm for attitude research. *Personality and Social Psychology Review, 14*(4), 351–367. https://doi.org/10.1177/1088868310366452

Knijnenburg, B. P., Reijmer, N. J., & Willemsen, M. C. (2011). Each to his own: How different users call for different interaction methods in recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems* (pp. 141–148). Association for computing machinery. 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., & Broeders, R. (2014). Smart sustainability through system satisfaction: Tailored preference elicitation for energy-saving recommenders. In *20th Americas conference on information systems* (AMCIS 2014) (pp. 1–15). AIS/ICIS.

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

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

Koren, Y., Rendle, S., & Bell, R. (2021). Advances in collaborative filtering. *Recommender systems handbook* (pp. 91–142). https://doi.org/10.1007/978-1-4899-7637-6_3

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. https://doi.org/10.1088/1755-1315/40/1/012087

Lika, B., Kolomvatsos, K., & Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications, 41*(4), 2065–2073. https://doi.org/10.1016/j.eswa.2013.09.005

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

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

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

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

Oonk, A. S. (2023) *Sustainable Goals: Energy-Saving Goals and Signposting in a Rasch-Based Recommender System* [Master's thesis, Eindhoven University of Technology, Eindhoven, Netherlands]. Retrieved August, 31, 2023. https://pure.tue.nl/ws/portalfiles/portal/306950304/Master_Thesis_Report_Aleid_Oonk.pdf

Schultz, P. W. (2014). Strategies for promoting proenvironmental behavior: Lots of tools but few instructions. *European*

Psychologist, 19(2), 107–117. https://doi.org/10.1027/1016-9040/a000163

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., 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

Starke, A. D., Willemsen, M. C. (2024). Psychologically informed design of energy recommender systems: Are nudges still effective in tailored choice environments?. In B. Ferwerda, M. Graus, P. Germanakos, M. Tkalčič (Eds.), *A Human-Centered Perspective of Intelligent Personalized Environments and Systems. Human–Computer Interaction Series*. Springer. https://doi.org/10.1007/978-3-031-55109-3_9

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

Urban, J., & Ščasný, M. (2012). Exploring domestic energy-saving: The role of environmental concern and background variables. *Energy Policy, 47*, 69–80. https://doi.org/10.1016/j.enpol.2012.04.018

Van den Broek, K., Bolderdijk, J. W., & Steg, L. (2017). Individual differences in values determine the relative persuasiveness of biospheric, economic and combined appeals. *Journal of Environmental Psychology, 53*, 145–156. https://doi.org/10.1016/j.jenvp.2017.07.009

Warren, C., Becken, S., & Coghlan, A. (2017). Using persuasive communication to co-create behavioural change – engaging with guests to save resources at tourist accommodation facilities. *Journal of Sustainable Tourism, 25*(7), 935–954. https://doi.org/10.1080/09669582.2016.1247849

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

Zangerle, E., & Bauer, C. (2022). Evaluating Recommender Systems: Survey and Framework. *ACM Computing Surveys, 55*(8), 1–38. https://doi.org/10.1145/3556536

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.