Beyond “one-size-fits-all” platforms

Beyond “one-size-fits-all” platforms: Applying Campbell’s Paradigm to test personalized energy advice in the Netherlands

A.D. Starke (a.d.starke@tue.nl)*, M.C. Willemsen*, C.C.P. Snijders*

*Human-Technology Interaction Group, Eindhoven University of Technology

Abstract

When analyzing ways in which people save energy, most researchers and policy makers conceptually differentiate between curtailment (e.g. unplugging chargers) and efficiency measures (e.g. installing PV cells). However, such a two-dimensional approach is suboptimal from both a conceptual and policy perspective, as it does not consider individual differences that determine energy-saving behavior. We propose a different, one-dimensional approach, applying Campbell’s Paradigm through the Rasch model, in which both curtailment and efficiency measures are intermixed on a single scale and ordered according to their behavioral costs. By matching these behavioral costs to individual energy-saving attitudes, we investigate to what extent attitude-tailored energy-saving advice can help consumers to save energy.

We present the results of two studies. The first study (N=263) reliably calibrated a one-dimensional Rasch scale that consists of 79 energy-saving measures, suitable for advice. The second study employed this scale to investigate how users (N=196) evaluate attitude-tailored energy-saving advice in a web-based energy recommender system. Results indicate that Rasch-based recommendations can be used to effectively tailor energy-saving advice and that such attitude-tailored advice is more adequate than a number of non-personalized approaches.

Keywords

Conservation advice, energy efficiency, Rasch Model, recommender systems

1. Introduction

When discussing energy policy, many authors conceptually differentiate between two types of energy-saving measures: *efficiency* and *curtailment* [1–3]. In most studies, *efficiency* comprises one-time investments in home equipment, such as installing double-glazed windows, while *curtailment* involves repetitive changes in energy-related behaviors, such as turning off the lights after leaving a room [4–6].

The curtailment-efficiency dichotomy is often used to point out a discrepancy between what conservation researchers consider effective energy-saving measures and what the public perceives as such [7]. Although efficiency measures tend to yield higher energy savings than curtailment [1,8], consumers still seem to apply more curtailment than efficiency in their households, perceiving curtailment measures as easier to implement and perhaps also as more effective [4,9,10]. Most policy makers attempt to overcome this gap by educating consumers on energy conservation [11], predominantly promoting the use of state-of-the-art efficiency measures [2,12,13]. However, we argue that using such an approach to energy-saving behavior might be suboptimal, both from a policy and a conceptual point of view.

From a policy perspective, merely informing consumers about energy conservation has had little impact on their energy-saving behavior [14,15]. This lack of effectiveness may be due to policy makers often recommending the same energy-saving measures to different consumers [16]. Such a one-size-fits-all approach fails to consider interpersonal differences that determine energy-saving behavior [11], including one's energy-saving attitude or knowledge level [17,18].

Furthermore, the conservation literature also contests a curtailment-efficiency dichotomy on a conceptual level [19]. Multiple studies have argued either that conservation behavior needs to be analyzed using more than two underlying dimensions [1,5,20], or that a single dimension is sufficient [21,22]. Evidence for the latter was delivered by Urban and Ščasný [23], who showed through the psychometric Rasch model that curtailment and efficiency measures tap into the same dimension [24,25], and outperform a two-dimensional approach in terms of model fit. This one-dimensional trait manifests itself as a two-sided measurement scale [22,23], on which on the one hand energy-saving measures are ordered on their behavioral costs or execution difficulty, and on the other hand persons are ordered on their energy-saving attitude.

Studies that examine the appropriateness of Rasch scales in the sustainability domain tend to focus on conceptual issues, such as how to define an energy-saving attitude [22], but have hardly discussed policy implications. We consider this a missed opportunity, as a Rasch scale of energy-saving measures, along with the model's formal relation between measures and persons [24], allows policy makers to consider attitudinal differences between individuals in energy-saving advice. For example, such a scale could be employed to estimate an individual's energy-saving attitude and subsequently recommend appropriate energy-saving measures for that attitude. This way, energy policy makers could move beyond the often employed but rather ineffective one-size-fits-all strategies [4,26–28]. In fact, such tailored interventions have shown to be more effective in persuading consumers to adopt energy-saving measures than nondiscriminatory approaches, as well as achieve higher energy savings [3,17,29,30].

To be able to provide tailored advice based on a Rasch scale, two important extensions to earlier work are required. First, in order to have sufficient measures at our disposal to recommend novel measures to consumers, our scale requires a much larger number of energy-saving measures than those used in earlier studies [23,31]. Second, we test whether measuring a small subset of scale measures per person is sufficient to estimate a person's attitude and tailor advice towards it. Eventually, these extensions allow us to evaluate the effectiveness of Rasch-based advice compared to other advice strategies, such as one-size-fits-all approaches.

In the upcoming sections, we first provide some background of the Rasch model by introducing 'Campbell's Paradigm'. Then, we explain how we use the Rasch model to provide energy-saving advice using a recommender system. Finally, we present two studies: a calibration study in which we construct a Rasch scale of energy-saving measures, followed by a user study that investigates how tailored energy-saving recommendations can be created and how they are evaluated.

2. Theory

Usage of the Rasch model in the current context follows the logic of Kaiser et al. [22], who have presented the Rasch model as an alternative way to conceptualize and measure a person's attitude, a topic heavily debated by social psychologists [32–34].

2.1. Campbell's Paradigm

Kaiser et al. [22] propose an attitude theory that is named after Donald Campbell [33]. Rather than using evaluative statements, as in conventional attitude research [cf. 35], Campbell's Paradigm draws upon a wider range of responses that intermixes both behavioral self-reports and intentions. It describes an axiomatic connection between a person's attitude towards a certain behavioral goal and the behaviors that person is willing to engage in to achieve that particular goal [31,36,37]. In the conservation context, Campbell's Paradigm postulates that one's attitude for saving energy becomes apparent through the different energy-saving measures a person is willing take [22].

Energy-saving measures form a specific latent class pertaining to the goal of saving energy [23], as long as these measures differ in their execution difficulty [22]. This execution difficulty is operationalized as a behavioral cost level [22], which comprises costs in terms of, among others, cognition, money and time [38], and can differ substantially between measures. For instance, verbal statements typically come with low behavioral costs (e.g. stating that saving energy is important), while actually installing energy-efficient appliances come with much higher costs (e.g. installing solar PV on one's rooftop requires costs in terms of cognitive effort, money, and time).

Furthermore, Campbell's Paradigm prescribes that persons committed to performing an energy-saving measure carrying high behavioral costs are also likely to perform measures with lower costs [22,36]. For example, an individual who applies thermal insulation, which is a relatively high-cost measure, is also likely to turn off the lights after leaving a room, which is a relatively low-cost measure [23]. If an individual discloses engagement levels for multiple energy-saving measures, it becomes possible to estimate that individual's energy-saving attitude. In turn, one's attitude can predict what other measures are likely to be performed.

An important assumption of Campbell's Paradigm is that these behavioral cost levels hold for each individual and can predict behavior accordingly. However, Kaiser et al. [22] stress that performing a measure with high behavioral costs does not deterministically guarantee the execution of those with lower costs, as there may be contextual irregularities. Campbell's Paradigm accounts for these irregularities through the Rasch model, which formalizes the axiomatic relationship between attitudes and behavioral costs in a probabilistic model.

2.2. The Rasch Model

Commonly used in psychometrics and based on item-response theory [cf. 39], the Rasch model predicts whether a person performs a particular item or behavior. It does so by modelling a latent trait as a function of the behavioral difficulties of a set of trait items [24]. In an energy-saving context, Rasch captures energy-saving measures and persons onto a one-dimensional measurement scale [22,23]. The level of behavioral costs of a measure is proportional to the number of individuals in a sample performing it, where commonly performed measures yield lower behavioral costs than those performed less frequently. Conversely, the strength of a person's attitude increases proportionally with the number of performed scale measures.

The Rasch model used in the current study predicts whether an individual n performs a measure i or not, as a logistic function of the arithmetic difference between that individual's attitude θ and the measure's behavioral costs δ. Equation 1 shows how an increase in one's energy-saving attitude increases the probability that one performs a certain energy-saving measure [cf. 40]. Moreover, the current study assumes that such a probability distribution, referred to as an item-characteristic curve [24], is similarly-shaped for each scale item.

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

Figure 1 depicts the Item-Characteristic Curves (ICC's) of three different energy-saving measures, illustrating how the engagement probability (y-axis), attitude (x-axis), and behavioral cost parameters (the curves) relate to each other. The energy-saving attitude θ, expressed in logistic scale units (logits), varies between individuals and leads to different engagement probabilities for the various energy-saving measures. In addition, the behavioral cost level δ of a measure is equal to the 50%-engagement probability point of an ICC. Hence, if a person's energy-saving attitude is equal to the behavioral cost level of an energy-saving measure, then that person has a 50% probability of performing that measure. Consider for example a person with an attitude θ of 0.9 logits: Figure 1 shows that this person has a 50% probability to use a water-saving showerhead (δ = 0.9 logit), but would be much more likely to own double-glazed windows (δ = −1.72 logit), namely 93%.

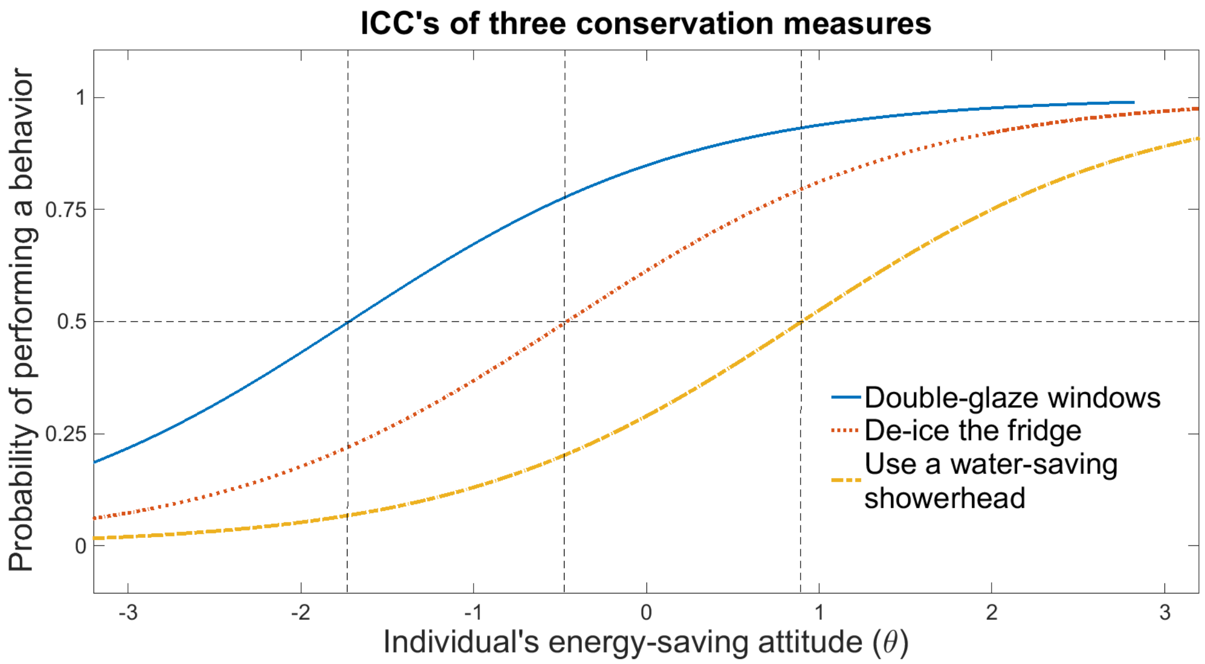

Fig. 1. Item-characteristic curves (ICC's) of three conservation measures with different behavioral cost levels, yielding varying engagement probabilities as a function of an individual's energy-saving attitude.

2.3. Dimensionality of energy-saving behavior

The adequacy of using the Rasch model to describe energy saving and other environmental behaviors finds empirical support in multiple studies [21,41]. In particular, Urban and Ščasný [23] show through a survey on self-reported energy-saving measures in ten OECD countries that the Rasch model outperforms approaches that differentiate between curtailment and efficiency, detailing two main findings. First, a two-factor model does not properly represent patterns of energy-saving behavior, while the one-dimensional Rasch model does have an acceptable model fit. Second, although the average behavioral cost level of efficiency measures exceeds that of their efficiency counterparts, both curtailment and efficiency measures are intermixed as part of the unidimensional Rasch construct. For instance, in some countries, an efficiency measure as the use of green energy had a lower behavioral cost level than turning off the lights after leaving a room, a curtailment measure.

2.4. Attitude-tailored advice

To determine how to use the Rasch model to provide suitable attitude-tailored advice to individuals, we can draw upon several model characteristics. We have already shown that, regardless of one's energy-saving attitude, some measures bear lower behavioral costs than others and are therefore more likely to be performed. In addition, such a high probability implies that a measure's benefits outweigh its behavioral costs [22], deeming such low-cost behaviors to be more feasible than their high-cost counterparts. However, only suggesting low-cost measures to an individual would lead to rather redundant advice, as most of them would already be performed.

To account for this trade-off between feasibility and redundancy, we expect that tailoring energy-saving advice around an individual's energy-saving attitude is an effective approach. This would result in adequate advice, as a measure's perceived benefits would be roughly equal to its behavioral costs, while such measures could still be relevant or novel to an individual as they would still have about a 50% probability of not being performed yet (see Figure 1).

However, it is not clear at which engagement probability level such energy-saving advice would the most effective. For example, if one would reduce the behavioral cost level of a suggested measure to 1 logit below an individual's attitude, it would lead to an increase in the recommendation's feasibility at the expense of novelty, due to an increased engagement probability (from 50% to 75%; see Figure 1). On the other hand, an increase of 1 logit in behavioral costs might be relatively unfeasible, but would lead to more novel measures due to an engagement probability of only 25%. By comparing multiple attitude-cost differences around an individual's attitude this way, the current research investigates at which disposition an optimal trade-off between advice feasibility and novelty is achieved.

2.5. Energy recommender systems

In order to investigate how individuals evaluate Rasch-based advice, we need to be able to estimate an individual's energy-saving attitude and advise scale measures accordingly. A similar setup involving adaptive advice is used in recommender systems [42], which are information filtering systems that rely on a user's input, such as ratings or self-reported behavior, and subsequently provide user-tailored recommendations by drawing upon an extensive database of domain items, such as movies or online dates [43,44]. Analogously, energy recommender systems can help consumers to overcome their often limited knowledge on what actions to take to reduce their energy consumption [26,45], allowing them to discover effective energy-saving measures that fit their profile [30]. However, prior research on energy recommender systems has predominantly focused on optimizing the user experience by matching the preference elicitation method with the domain knowledge of the user [30]. Similarly, the Rasch scale developed in the current research has already been applied in another energy recommender study that focuses on user experience aspects [46], which was published ahead of the current paper. However, the present paper discusses more in-depth the theoretical and policy implications of the Rasch model.

The current research examines the validity of a one-dimensional Rasch construct for energy-saving behavior, as well as how it can be used for attitude-tailored advice. While preceding research inferred constructs that captured at most 40 environmental behaviors [41], we design a web-based conservation tool that draws upon a database of 88 energy-saving measures [47], which we use to fit a Rasch scale that is suitable for an advice context. To allow for such advice, the system asks its users to indicate whether they already perform various energy-saving measures and subsequently assigns the user a Rasch-based attitude, to which it adapts the presented advice. Although most Rasch studies in the ecological domain have estimated attitudes using all scale items, the Rasch model is capable of making reliable parameter estimates when only presenting a subset of measures to each user [48], which is a common approach in recommender system studies [cf. 44].

2.6. Research setup

We present two studies to examine how individuals evaluate Rasch-based, attitude-tailored advice. In Study 1, we examine and affirm that it is possible to reliably construct a one-dimensional Rasch scale that consists of a large number of energy-saving measures, while only presenting a subset of these measures to each user. Moreover, we confirm earlier findings of Urban and Ščasný [23], testing whether curtailment and efficiency measures are also part of a single dimension when using a larger and more diverse set of energy-saving measures.

In Study 2, we apply the scale calibrated in the first study to a basic web-based recommender system, examining two separate matters. First, we examine to what extent the attractiveness of a measure depends on its behavioral cost level, using the difference between attitude and behavioral costs as a key predictor for how advice is evaluated. For this purpose, we estimate each user's energy-saving attitude and subsequently explore the feasibility-redundancy tradeoff by recommending measures at multiple levels of attitude cost-difference (i.e. engagement probabilities), which either have a relatively high, low, or equal behavioral cost level compared to a user's attitude.

Second, this setup allows us to investigate whether tailored energy-saving advice has benefits over a non-personalized approach. Specifically, we examine how advice should be tailored towards a user's energy-saving attitude to maximize the probability that it is adequate for that user, while comparing it to other non-personalized approaches. For different attitude-cost dispositions, we compare the predicted attractiveness of a measure to the probability that a measure is already performed by the user. Although we expect users to perceive measures with a low behavioral cost level as feasible and attractive (cf. section 2.4), the high probabilities that such measures are already performed might deem advice with somewhat higher behavioral costs, tailored around a user's attitude, to be the most adequate (cf. section 2.2).

3. Study 1: Constructing a unidimensional Rasch scale

3.1. Method

3.1.1. Participants and procedure

Study 1 investigated whether we can construct a one-dimensional scale that consists of a large number of energy-saving measures, using only a limited number of questions. In total, we recruited 263 Dutch-speaking participants (57% male) on multiple social media websites, among which we raffled five gift vouchers of €10 each. Participants interacted with an energy saving web-tool, which contained 88 energy-saving measures of both the curtailment and efficiency type, adapted from Knijnenburg and Willemsen [47], such as ‘turn off lights after leaving a room’ or ‘install solar PV’.

To assess each participant’s energy-saving attitude, we divided our database of 88 measures into thirteen subsets, based on how frequently they were chosen in the recommender system of Knijnenburg and Willemsen [47]. Subsequently, we randomly sampled one measure from each subset to present to the participant, thirteen in total. To each measure, presented one at a time, participants could either respond ‘yes’, ‘no’, or ‘not applicable’ (N/A). ‘N/A’ responses were expected for measures for which a participant’s housing situation was not practical. For instance, a participant who did not have a garden would respond ‘N/A’ to a measure about efficient garden lighting.

3.2. Results

We collected responses from each participant to thirteen measures and analyzed the resulting item-person matrix using Winsteps software [49]. From this, we successfully inferred a one-dimensional scale for energy-saving behavior, which followed standard fit guidelines by achieving medium to high levels of reliability [24].

3.2.1. Misfit analysis

We performed a standard misfit analysis on the presented energy-saving measures, using two cut-off criteria. First, measures were required to have a high applicability rate (i.e. a low N/A-ratio) to effectively determine a participant's energy-saving attitude. Second, we analyzed whether measures fitted the one-dimensional construct, checking for either 'underfit' or 'overfit' problems to ensure the scale quality [cf. 24]. Underfit occurs when a response pattern is inconsistent with one's attitude or a measure's behavioral cost level [48]. In contrast, a measure was marked as overfit if it fitted too perfectly, either varying too little among user responses or signaling the presence of redundant measures in a Rasch response set [48,50], for instance, if a scale contained multiple measures with similar cost levels.

We excluded nine energy-saving measures and three participants from our analysis. These participants had either given more than 50% 'does not apply' responses, were identified as serious misfits using Winsteps software [49], or both. Misfitted measures and participants violated mean square and standardized fit guidelines [24,48]. The remaining 79 energy-saving measures, along with their infit indices, are listed in Table 1 and met the prescribed infit guidelines by not exceeding a mean square of 1.4. A few measures were possibly overfit, as the underlying data seemed too predictable (they had a relatively low standardized value (ZSTD < −2.0)) [24]. Although this could suggest that additional dimensions were present in the data, we found no evidence for this (cf. section 3.2.3).

The measures were used in a Rasch analysis that drew upon dichotomous responses ('yes' or 'no'). We considered N/A-responses as 'missing', because some measures suggested a behavior that a user might not have been able to perform in the first place. In our view, most of these responses were not representative for the strength of a user's attitude. In addition, we checked whether considering N/A-responses as 'no' altered the scale's reliability, but this merely led to a small decrease in person reliability without any other differences.

3.2.2. Model reliability

Table 1 details the behavioral cost levels (*δ*) of scale measures in study 1, along with those for study 2 which will be discussed later. The *δ*-levels for study 1 ranged from −5.73 to 5.49 (M = 0.06; SD = 2.14), encompassing all person attitudes present on the scale (Min: −3.45 to 4.31; M = 0.03 logits; SD = 1.20). We inferred model validity through a number of indicators. First, the model's measures had a high item reliability α of 0.92, suggesting that the scale's order in terms of behavioral costs is likely to be reproducible [40]. Second, the model had a large item separation of 3.48 and an item mean square close to the identity value (0.98), both indicating that the scale is suitable for measurement [48]. The inferred participant attitudes had a person reliability of 0.59 and a separation of 1.14, which were acceptable values [24], particularly because we only used thirteen measures per person. In addition, the model's fit statistics were comparable to studies which employed fewer items [cf. 23].

Table 1

Behavioral cost levels (δ) and infit indices (Mean Square (MS) and Standardized Values (ZSTD)) for two Rasch scales of energy-saving measures, for Study 1 and Study 2 (cf. section 4.2.1). The two scales have similar behavioral cost levels: $r(79) = 0.88$. 'Set' denotes the 13 different sets from which measures are sampled for attitude calibration in Study 2.

		Study 1			**Study 2**			
			Infit			Infit		
#	Name of energy-saving measure (translated)	**δ**	MS	ZSTD	**δ**	MS	ZSTD	Set
1	Save up laundry	**−5.73**	Min	Min	**−3.23**	1.14	0.4	1
2	Take a shower instead of a bath	**−4.82**	0.95	0.2	**−4.41**	Min	Min	1
3	Wash laundry at low temperatures	**−3.95**	1.14	0.4	**−1.64**	1.01	0.1	1
4	Air-dry laundry	**−3.69**	0.99	0.1	**−2.93**	1.19	0.5	1
5	Use a laptop instead of a desktop PC	**−3.45**	1.04	0.2	**−3.62**	1.16	0.5	1
6	Turn off the lights after leaving a room	**−2.97**	0.85	−0.3	**−2.78**	0.69	−1.0	1
7	Use public transportation instead of a car	**−2.90**	1.10	0.4	**−2.52**	1.42	1.4	1
8	Use a woolen blanket instead of an electric blanket	**−2.51**	0.99	0.1	**−3.03**	1.05	0.3	2
9	Use properly sized cooking equipment	**−2.51**	0.88	−0.2	**−2.69**	0.66	−1.2	2
10	Lower the thermostat while away from home	**−2.49**	0.97	0	**−1.92**	0.79	−1.0	2
11	Do not put warm things in the fridge	**−2.45**	0.69	−1.3	**−2.40**	0.97	−0.1	2
12	Turn off the PC screen after use	**−2.20**	0.97	0	**−0.71**	1.08	0.6	2
13	Close the curtains/shutters in the evening	**−2.09**	1.07	0.4	**−1.57**	1.03	0.3	2
14	Shift gears at low speeds	**−1.89**	1.02	0.2	**−2.07**	0.81	−1.1	3
15	Cook with a lid on the pan	**−1.81**	0.91	−0.5	**−1.81**	0.93	−0.5	3
16	Use energy-saving bulbs (CFL's)	**−1.75**	0.81	−1.1	**−1.31**	0.76	−2.7	3
17	Double-glaze windows	**−1.72**	0.81	−0.9	**−1.24**	1.01	0.1	3
18	Air rooms for 20 minutes daily	**−1.51**	0.95	−0.3	**−1.21**	1.13	1.2	3
19	Cook on gas stove instead of electric	**−1.36**	1.34	1.7	**−2.23**	1.30	1.6	3
20	Lower the thermostat one degree	**−1.25**	1.02	0.2	**−0.47**	1.01	0.1	4
21	Set thermostat to 14ºC before going to bed	**−1.20**	0.97	−0.2	**−0.90**	1.14	1.4	4
22	Do not defrost food using a microwave	**−1.18**	1.04	0.3	**−0.52**	1.16	1.7	4

23	Turn off the TV instead of stand−by	**−1.06**	1.26	1.1	**−0.74**	0.96	−0.4	4
24	Maintain correct tire pressure	**−0.94**	0.86	−0.5	**−0.27**	1.06	0.5	4
25	Stir-fry food	**−0.91**	1.08	0.7	**−0.62**	1.01	0.1	4
26	Turn off the PC at the main switch	**−0.81**	0.90	−0.9	**−0.21**	0.99	−0.1	5
27	Turn off the coffee machine completely	**−0.57**	0.91	−0.7	**−1.30**	0.94	−0.4	5
28	Turn off the dishwasher after use	**−0.57**	0.77	−1.8	**−0.37**	0.99	−0.1	5
29	Insulate the cavity wall	**−0.51**	1.20	0.9	**0.50**	0.89	−0.9	5
30	Turn off the washing machine completely	**−0.49**	0.91	−1.0	**−0.22**	1.00	0	5
31	De-ice the fridge	**−0.46**	0.96	−0.3	**0.59**	0.98	−0.2	5
32	Unplug chargers	**−0.32**	1.01	0.1	**−0.44**	0.98	−0.2	6
33	Take short showers	**−0.29**	1.13	1.6	**0.52**	0.92	−0.8	6
34	Hand-wash dishes (no dish washer)	**−0.22**	1.33	2.7	**−0.72**	1.32	2.9	6
35	Configure PC power management	**0.00**	1.15	1.4	**0.29**	1.00	0.1	6
36	Shorten PC/laptop stand-by time	**0.01**	0.90	−0.6	**−0.43**	0.84	−1.8	6
37	Air clothes instead of washing them	**0.07**	1.06	0.7	**−0.39**	1.10	1.1	6
38	Clean the cooker hood suction filters	**0.14**	1.02	0.2	**0.29**	1.10	0.7	7
39	Place fridge in a suitable position	**0.18**	1.00	0.1	**−0.14**	0.78	−1.2	7
40	Use LED lighting	**0.37**	1.04	0.5	**0.38**	0.89	−0.8	7
41	Decalcify your coffee machine and/or kettle	**0.40**	0.96	−0.3	**0.35**	0.88	−1.1	7
42	Sweep instead of using a vacuum cleaner	**0.43**	1.21	1.6	**0.40**	1.06	0.6	7
43	Use a smart thermostat	**0.47**	1.06	0.5	**−0.11**	0.88	−1.4	7
44	Put a weather strip on the door	**0.47**	0.75	−2.5	**0.19**	0.89	−1.4	8
45	Use a HE boiler or CHP	**0.47**	0.96	−0.2	**0.98**	0.96	−0.2	8
46	Use household devices without displays	**0.48**	1.20	1.3	**2.67**	0.99	0.1	8
47	Use an ‘A+’ energy-class fridge	**0.51**	0.87	−1.3	**−0.09**	0.90	−1.0	8
48	Install motion sensors	**0.51**	0.81	−1.9	**−0.02**	1.06	0.5	8
49	Insulate floors	**0.56**	0.91	−0.5	**0.29**	0.99	0	8
50	Use a mini PC instead of desktop computer	**0.59**	0.89	−0.9	**3.89**	Max	Max	9
51	Make coffee without using a heating plate	**0.74**	0.94	−0.5	**−0.84**	0.99	0	9
52	Decalcify the washing machine	**0.75**	1.15	1.1	**0.59**	1.06	0.4	9
53	Use green power	**0.85**	1.19	1.2	**0.22**	0.93	−0.6	9
54	Turn off the fridge while on holiday	**0.87**	1.11	0.8	**1.84**	1.11	0.5	9
55	Turn off the PC when away from keyboard	**0.88**	1.17	1.0	**−0.74**	0.82	−1.5	9

56	Use a water-saving showerhead	**0.90**	1.03	0.3	**0.34**	1.09	0.8	10
57	Put your shirts briefly in the laundry dryer instead of ironing them	**0.96**	1.00	0.1	**1.15**	0.83	−0.7	10
58	Cover the windscreen of your car	**0.96**	0.82	−0.4	**0.81**	0.95	−0.1	10
59	Replace dimmer switches	**0.99**	0.73	−1.5	**1.04**	1.35	1.7	10
60	Use an ‘A-label’ energy-saving laundry dryer with a heat pump	**1.17**	1.05	0.3	**1.36**	0.96	−0.2	10
61	Use day and night tariffs	**1.21**	0.90	−0.6	**0.75**	0.91	−0.8	10
62	Set boiler temperature to 65 degrees Celsius	**1.24**	1.11	0.7	**1.04**	0.95	−0.3	11
63	Set the mixing valve at a lower temperature	**1.31**	0.82	−1.0	**1.17**	0.90	−0.7	11
64	Put weather strips on the windows	**1.46**	0.66	−2.1	**0.59**	0.94	−0.6	11
65	Insulate hot water pipes	**1.53**	0.85	−0.7	**0.79**	1.03	0.3	11
66	Clean the water heater	**1.63**	0.88	−0.5	**1.26**	1.00	0.1	11
67	Install a door closer	**2.34**	1.39	1.2	**1.39**	0.84	−0.7	11
68	Turn off the oven before the end of cooking time	**2.40**	0.88	−0.4	**2.27**	1.13	0.5	12
69	Keep the rear of the fridge dust-free	**2.43**	0.75	−1.0	**1.25**	0.93	−0.3	12
70	Apply heat reflection foil to radiators	**2.77**	1.07	0.3	**3.46**	1.41	0.7	12
71	Replace a radio alarm with a ‘classic’, un-plugged alarm clock	**2.88**	1.06	0.3	**3.51**	1.39	0.7	12
72	Use a cabled telephone instead of a handheld phone	**2.91**	0.99	0.2	**2.49**	1.19	0.5	12
73	Install solar PV	**3.17**	1.36	0.9	**1.60**	0.76	−0.5	12
74	Install a solar boiler	**3.26**	1.10	0.4	**2.53**	1.06	0.3	13
75	Slow down the PC processor	**3.70**	1.22	0.6	**3.47**	0.96	0.3	13
76	Use a pull bell instead of an electrical bell	**3.82**	0.82	−0.1	**3.20**	1.05	0.4	13
77	Wash using a ‘hot-fill’ washing machine	**4.04**	0.87	0.1	**2.36**	0.91	0	13
78	Use software for dynamic energy use in a laptop or PC	**5.18**	Max	Max	**0.52**	0.94	−0.2	13
79	Erect a small wind mill to produce electric energy	**5.49**	Max	Max	**4.42**	Max	Max	13

3.2.3. Dimensionality

We investigated the dimensionality of participants' responses by analyzing the explained variance for both one- and multidimensional constructs, consistent with dimensionality checks in other domains [24]. Using Winsteps software [49], we determined that the one-dimensional construct explained 39.6% of the variance in the data, slightly overfitting the predicted model fit (39.1%). The remaining 60.9% was quantification variance, which is residual variance caused by the Rasch model's estimated probabilities for discrete, dichotomous events (0 or 1).

We checked whether multiple scale measures shared an unexpected response pattern in the residual variance and formed an additional dimension [cf. 51]. A principal component analysis on the standardized residuals showed little evidence to expect more than one dimension in our sample, as an additional factor would only result in a trivial increase of 1.8% in the proportion of explained variance. Moreover, measures in these additional dimensions did not seem to form meaningful factors, which further confirmed earlier research that an individual's energy-saving attitude could be assessed on a one-dimensional scale [22,23].

Figure 2 depicts the distribution of curtailment and efficiency measures across the scale, in terms of their behavioral costs. As in Urban and Ščasný [23], curtailment and efficiency measures formed a single construct but were not uniformly distributed across the one-dimensional scale. This is illustrated by a higher density of efficiency measures on the high-cost end of the scale: curtailment measures (N=46; M_{cur} = −0.67 logistic units) had on average lower behavioral costs than efficiency measures (N=33; M_{eff} = 1.08 logistic units), consistent with findings that efficiency measures are on average less likely to be performed than curtailment [7]. However, Table 1 points out that some curtailment measures still had particularly high behavioral costs (e.g. keeping the rear of a refrigerator dust-free, δ = 2.43), while some efficiency measures had very low behavioral costs (e.g. using a laptop instead of a PC, δ = −3.45).

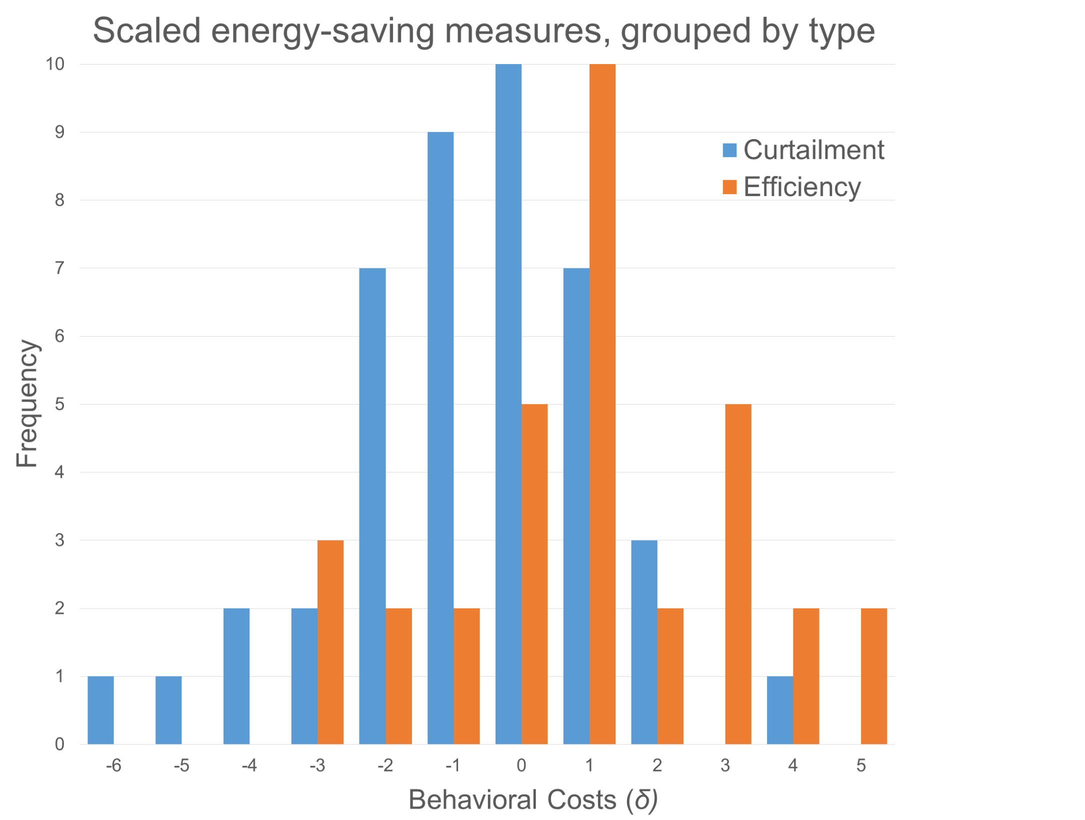

Fig. 2. Distribution of curtailment and efficiency measures fitted on the Rasch scale, relative to their behavioral cost level, confirming that efficiency measures have on average higher behavioral costs, but are intermixed on the scale.

3.3. Conclusion

Study 1 delivered a unidimensional scale of 79 energy-saving measures, suitable for advising a heterogeneous group of persons due its diversity in behavioral cost levels. Consistent with Urban and Ščasný [23], we have found that energy-saving behavior should be described as a one-dimensional rather than a two-dimensional construct, with curtailment and efficiency measures distributed across the entire Rasch scale. The lower average behavioral costs of curtailment measures, compared to their efficiency counterparts, is consistent with earlier findings that consumers are more likely to perform curtailment measures rather than efficiency [9].

Moreover, our study sets forth a behavioral explanation for the (non-)performance of energy-saving measures. Since curtailment and efficiency measures overlap on our behavioral cost scale, a simple dichotomy between those two dimensions is unable to precisely predict whether a measure will be performed or not, much unlike the formal Rasch model. This supports claims in earlier research that the curtailment-efficiency dichotomy lacks empirical validity [5].

The moderate (for persons) and high (for items) scale reliability levels suggest that our estimation procedure has been appropriate. Using only 13 out of 79 scale measures to estimate a person's attitude does not compromise the scale quality in terms of item reliability, and also provides accurate and useful attitude estimates. Moreover, it sets forth a practical approach for applying a Rasch construct in a recommender system, as the attitude estimation procedure is short (in terms of time), as well as restricted (in terms of used measures), which allows multiple scale measures to be presented as novel recommendations to system users. We will explore such tailored energy-saving advice in study 2.

4. Study 2: Rasch-based energy recommendations

In our second study, we implemented the scale of 79 energy-saving measures, calibrated in Study 1, in a web-based energy recommender system to provide attitude-tailored advice. We investigated users' preferences for measures with different behavioral cost levels, which were either above, below or equal to a user's attitude. Using these evaluations and the probabilities that persons would perform certain measures, we determined to what extent such attitude-tailored advice was more appropriate than a non-personalized approach, comparing four different scenarios of energy-saving advice.

4.1. Method

4.1.1. Participants

Our recommender web-tool was distributed among the members of the JFS participant database at the Eindhoven University of Technology, the Netherlands. In total, 196 participants, none of whom had participated in study 1, completed study 2 and were compensated by entering a raffle (a 20% chance of winning €15). Each participant was redirected to our Dutch website 'besparingshulp.nl' (i.e. 'saving aid'), where they were told they would receive energy-saving advice.

4.1.2. Procedure: attitude calibration

Before the system presented its lists of recommendations, the users had to complete a few steps. First, we estimated a user's energy-saving attitude using a setup similar to Study 1. Users were presented thirteen energy-saving measures (out of 79) for which they had to indicate whether they performed them, by either responding 'yes', 'no', or 'not applicable' (N/A). To do so, we divided the ordered measures into thirteen equally large subsets across the scale's entire behavioral cost range (see Table 1), and randomly sampled a measure from each subset. A user's attitude was estimated by tallying the number of 'yes' responses and using the average behavioral cost level of the corresponding scale subset. For instance, if a user had answered 'yes' four times, this amounted to the average behavioral cost level of the fourth subset (−1.09; see Table 1). In addition, we corrected the total number of 'yes'-responses if any N/A-responses were given, adding the product of the number of N/A-responses and the proportion of 'yes'-responses to the total sum and rounding this to the nearest subset integer.

This calibration approach was somewhat simpler than traditional methods with a-priori known behavioral cost levels (e.g. anchored maximum likelihood estimation [24]), for it assumed that the presented sets of thirteen measures were roughly similar in terms the presented behavioral cost levels. However, this method of tallying positive responses was much easier to implement in a web-tool, while we found that the resulting estimates still correlated strongly with those determined in a traditional method (cf. section 4.2.1).

4.1.3. Procedure: recommendation lists

The experiment proceeded with two further steps. First, in order to make users more aware of the fact that the system would present tailored advice, users were shown an energy-saving score based on their estimated attitude, which amounted to a value between 1 and 10. Second, following this score, users were presented two consecutive lists of nine randomly-ordered energy-saving measures.

Each recommendation list contained three different types of measures (3x3 measures), in terms of their behavioral costs. A list included three measures with behavioral costs that fell below the user's attitude and were likely to be performed (each had a probability of 75%, or −1 logit compared to the user's attitude), three measures whose behavioral costs were roughly equal to the user's attitude (i.e. a probability of 50%; +0 logit), and three measures with relatively high behavioral costs, which were unlikely to be performed (i.e. a probability of 25%; +1 logit above the user's attitude).

Figure 3 depicts part of such a recommendation list and shows how each measure depicted a name, such as 'Turning off the lights after leaving a room', and a short, explanatory description. For both recommendation lists, users had to perform two tasks. First, if users either already performed a measure or it was not applicable to them, they were required to click one of the corresponding boxes next to that measure ('yes' or 'N/A'), which removed the measure from the list and placed it at the bottom of the screen. Second, to estimate preferences for the non-performed measures with different behavioral cost levels, users had to rank the remaining measures in order of appropriateness, dragging the most suitable measures to the top of the list and less preferred alternatives to lower positions.

Fig. 3. A partial screenshot of a recommendation list (in Dutch). The user drags a measure to a different position to elicit his/her preference for that measure.

4.1.4. Descriptive statistics

Although 196 participants completed our experiment, some users could not complete the rank-order task, as they had removed too many energy-saving measures from a recommendation list. We found that 24.6% of lists were left empty and 4.1% of lists had a single measure remaining, for example, because users had indicated to already perform all list measures. As these lists had not been ranked in preference order, we excluded them from the analysis. The reduced sample comprised 152 participants, which had rank-ordered 1,361 energy-saving measures in 279 different lists.

We compared the full sample (N = 196) with the reduced sample (N = 152) to check for any differences in terms of their descriptive statistics. The 196 participants (52.69% female) in the full sample had a mean age of 31.3 years (SD = 15.5; Min = 19; Max = 77), while the reduced sample of 152 participants (51.6% female) was found to be significantly younger (M = 27.3; SD = 11.6), when compared to the removed sample of 44 participants: $t(194) = 6.89$, $p < 0.001$. This suggested that some older users might have misunderstood our somewhat unconventional drag-and-drop ranking interface in figure 3, for instance by assuming that they had to indicate their preferences through the clickable boxes.

4.2. Results

We investigated the effectiveness of energy-saving recommendations that used a Rasch-based attitudinal user model, performing three different analyses. First, we validated the scale used in the energy recommender system by performing a Rasch model analysis. Second, we analyzed two lists of nine energy-saving measures, rank-ordered in appropriateness order, to reveal that users perceived measures with behavioral costs just below their own attitude as the most appropriate. Third, based on users' preferences and whether measures were already performed by users, we compared four different scenarios of energy-saving advice and determined that a strategy which tailored measures around a user's attitude was more adequate than three other, non-personalized approaches.

4.2.1. Rasch model analysis

We checked whether the Rasch scale from Study 1 used in our recommender system was reproducible, by fitting a scale on the sample of Study 2. To assess which measures were already performed by each participant, we used both the responses provided in the attitude calibration, as well as the responses given in the rank-order task by the reduced sample of 152 participants. Using Winsteps software [49], we observed that measures and participants could be reliably scaled as a one-dimensional construct, where measures ranged from -4.41 to 4.42 in terms of behavioral costs (M = 0.05; SD = 1.80), and participants ranged from −3.58 to 2.92 in terms of their energy-saving attitude (M = −0.35; SD = 1.00). Item and person reliability were good to high (α = 0.88; α = 0.73, respectively), as were their separation indices (2.69 and 1.65, respectively). In line with Study 1, we found that both curtailment and efficiency measures were intermixed on our one-dimensional construct and did not form meaningful additional dimensions.

Table 1, section 3.2 lists the scale measures and their behavioral cost levels for the Study 1 and Study 2 calibrations, as well as their infit indices. Using a pairwise correlation analysis, we found that the behavioral costs levels for the Study 2 scale seemed similar to those in

Study 1: $r(79) = 0.88$, $p < 0.001$. Moreover, the infit indices suggested that the scale measures were productive for attitude measurement.

Finally, we checked the accuracy of the system-estimated user attitudes, based on the subsets designated by the Study 1 scale. We found these estimated attitudes to be strongly correlated with the attitudes determined in the current Rasch analysis: $r(152) = 0.79$, $p < 0.001$. This implied that the simple estimation procedure was sufficiently accurate for an energy recommender system. However, we did check to what extent any small differences between the two constructs impacted the results of our next analysis in 4.2.2, by considering attitude and behavioral cost levels from both the Study 1 and Study 2 scales.

4.2.2. User preferences for different energy-saving measures

We analyzed 279 recommendation lists of non-performed measures, each rank-ordered on their appropriateness, and investigated to what extent the final order depended on the user's attitude and a measure's behavioral cost level. We employed a rank-ordered logit model for analysis, a specific discrete choice model which is suitable for analyzing a list of rank-ordered alternatives [cf. 52,53]. Each model and its predictors sought to explain a list's appropriateness order and assumed that a ranked measure was preferred to a measure ranked below it (9 = top-ranked item, 8 = second item, etc.)

We analyzed four different rank-ordered logistic regression models, outlined in Table 2. Each model included the difference between the user's energy-saving attitude and a measure's behavioral cost level as a continuous predictor, which reflected a measure's relative ease of execution. To check whether using different Rasch constructs would affect any results, models 1.1 and 1.2 used the behavioral cost levels of Study 1 in conjunction with the current experiment's attitude estimates, while models 2.1 and 2.2 used the attitudes and behavioral cost levels as re-calculated from the data in study 2 (see section 4.2.1.).

In addition, models 1.2 and 2.2 both considered a set of four control predictors, which either concerned the recommendation task or underlying attributes of the energy-saving measures.

First, we controlled for the starting position of a measure prior to the rank-order task (9 = the top item, 1 = the bottom item), to check for position effects. Second, we examined to what extent underlying attributes played a role in user's preferences, by checking whether the rank-order was affected by either a measure being of the curtailment type or not, a measure's estimated annual savings in kWh, or a measure's estimated investment costs.

In line with the predictions of the Rasch model, Table 2 shows that measures with relatively low behavioral costs were ranked as more appropriate than their high-cost counterparts. All attitude-cost differences in Table 2 had a positive and significant effect on a measure's final rank-order position: $z(1361) = [5.92;6.70]$, $p < 0.001$, showing that if a user's energy-saving attitude exceeded a measure's behavioral cost level, it increased that measure's perceived appropriateness. This systemic preference for comparatively low-cost measures can serve as a starting point for providing tailored advice.

Table 2

Coefficients and standard errors for four rank-ordered, logistic regression models, clustered at the user level. Each model predicted the appropriateness rank-order of recommendation lists, with higher values for top-ranked items. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

	Model 1.1 *($\chi^2(1) = 43.07$)*		**Model 1.2** *($\chi^2(5) = 84.99$)*		**Model 2.1** *($\chi^2(1) = 38.89$)*		**Model 2.2** *($\chi^2(5) = 63.51$)*	
	Coef. (S.E.)	Z	Coef. (S.E.)	Z	Coef. (S.E.)	Z	Coef. (S.E.)	Z
Scale study 1								
Attitude-cost difference	.25 (.038)	6.56***	.28 (.042)	6.70***				
Scale study 2								
Attitude-cost difference					.26 (.042)	6.24***	.26 (.045)	5.92***
Starting position			.080 (.016)	4.96***			.083 (.017)	5.02***
Measure is curtailment			.12 (.090)	1.39			.17 (.090)	1.85
Savings in kWh / year			−.00027 (.000097)	−2.79**			−.00019 (.000095)	−2.05*
Investment costs			−.000095 (.000086)	−1.11			−.00011 (.000089)	−1.34

The control variables had mixed effects on the perceived appropriateness of an energy-saving measure, but none of them affected the effect of attitude-cost difference. Both models 1.2 and 2.2 show position effects, as measures which were ranked higher prior to the rank-order task, were also more likely to be eventually ranked at a higher position: $z(1361) = [4.96; 5.02]$, $p < 0.001$. In addition, we found that the annual savings in kWh negatively affected a measure's perceived appropriateness in both models: $z(1361) = [-2.79; -2.05]$, $p < 0.05$, suggesting that the savings in kWh, although not presented during the experiment, decreased a measure's attractiveness, which might be due to efficiency measures having relatively high kWh savings in conjunction with high behavioral costs. The other control variables, whether a measure was of the curtailment type and its investment costs, did not significantly affect a measure's rank-order position in both models.

4.2.3. Analysis of tailored advice

We observed that among a list of non-performed measures, those with relatively low behavioral costs were perceived as the most appropriate. However, only recommending low-cost measures would be an ill-fated strategy from a Rasch model point of view, due to the high likelihood that such measures are already performed. Instead, adequate energy-saving advice should strike the right balance between feasibility and novelty, in terms of the difference between the user's attitude and a measure's behavioral costs. We therefore investigated how energy-saving advice should be tailored to maximize the probability that it will be adopted, exploring the trade-off between feasibility (i.e. what users indicated to prefer) and novelty (i.e. the non-adoption probability that follows from the Rasch model). Subsequently, we also examined to what extent such attitude-tailored advice would be more adequate than three non-personalized approaches using different behavioral costs levels.

To determine the adequacy of attitude-tailored advice, we employed a scenario in which a single measure was recommended to a user. We approximated the probability that this measure would be attractive or feasible, by the probability it would be ranked first in our experiment's recommendation list. Since these recommendation lists only comprised non-performed measures, a recommendation's adequacy probability depended on the product between a measure's attractiveness, given that it was not yet performed, and the probability that it was not performed, calculated by the Rasch model. This led to Equation 2:

$$\mathrm{P(Recommendation\ is\ adequate)} = \mathrm{P(Attractive\ |\ Not\ yet\ performed)} * \left(1 - \frac{e^{\theta_n - \delta_i}}{1+e^{\theta_n - \delta_i}}\right) \qquad (2)$$

Using equation 2, we examined whether a certain attitude-cost disposition would be more adequate than any other. To approximate a measure's attractiveness, we conducted a post-estimation analysis on model 2.1 of the rank-ordered, logistic regression analysis (see Table 2). For a total of 279 recommendation lists and 1361 user-measure pairs, we inferred the probability that a non-performed measure was ranked first, given other measures in the list.

Figure 4 illustrates the results of our post-estimation analysis. In the left panel, we outlined the probability that a measure was attractive (i.e. ranked first) in red, P(Attractive | Not yet performed), and the probability that it was not performed by the user in blue (i.e. 1 minus the Rasch model), both as a function of the attitude-cost difference. Consistent with the results reported in section 4.2.2., we found that measures with comparatively low behavioral costs (i.e. a positive attitude-cost difference), were perceived as more attractive than those with higher behavioral costs. Moreover, the responses on whether measures were already performed were found to be consistent with the Rasch model (see Figure 1, section 2.2).

The right-hand side of Figure 4 depicts the product of these two probabilities, P(Recommendation is adequate), as described by equation 2. It shows that if a recommender system were to suggest a single measure to a user, it would be the most adequate around the attitude-cost disposition of −0.5, yielding probabilities of about 0.13. In contrast, recommending measures with comparatively low behavioral costs (i.e. a positive attitude-cost difference), led to smaller adequacy probabilities despite having a higher probability of being ranked first, reflecting a feasibility-redundancy tradeoff.

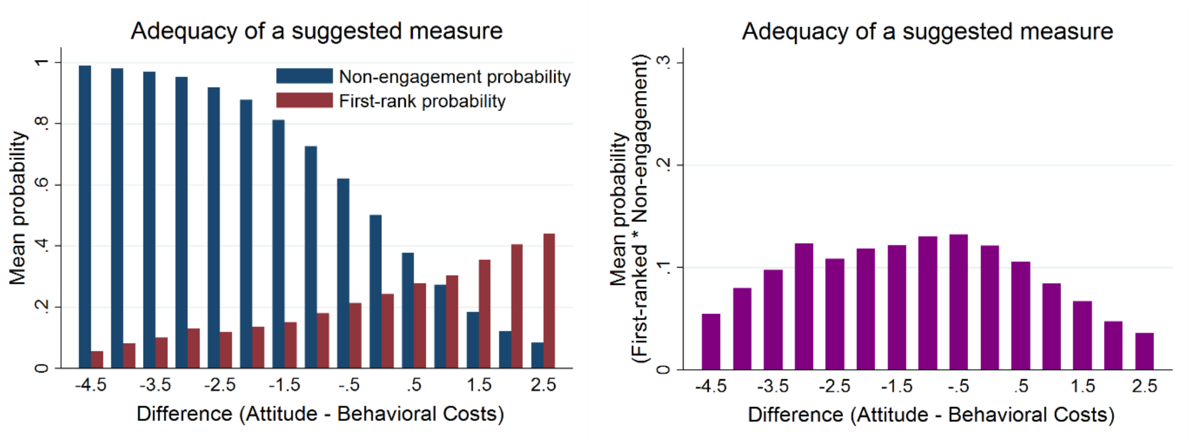

Fig. 4. Left panel: the probability that a measure is adequate for a user, based on its attractiveness (in red) and engagement level (in blue). Right panel: the product of the two probabilities representing the adequacy of the measure.

In addition, as Figure 4 has an optimum, this also suggests that recommending any measure randomly is less adequate than suggesting an attitude-tailored measure. To determine how large this difference actually is, we compared the adequacy of tailored advice with three different non-personalized approaches, each considering a scenario in which a single measure would be suggested to all users in our Study 2 sample. We compared a tailored measure, which formed the best match with a user's energy-saving attitude ($\theta - \delta = -0.5$ logits; 'tailored'), to three non-tailored measures: one measure with a low behavioral cost level ($\delta = -3$ logits; 'low'), one with average behavioral costs ($\delta = 0$ logits; 'average'), and one high-cost measure ($\delta = +3$ logits; 'high'). We used equation 2 to compute the adequacy of each approach, approximating a measure's attractiveness by fitting a simple linear regression model on the first-rank probability data in the left-hand chart of Figure 4. This led to the following model:

$$\mathrm{P(Advice\ adequacy)} = (0.054 * (\theta_n - \delta_i) + 0.24) * \left(1 - \frac{e^{\theta_n - \delta_i}}{1 + e^{\theta_n - \delta_i}}\right) \tag{3}$$

Figure 5 summarizes our estimation results, depicting the mean advice adequacy probabilities for the four different scenarios. We found that tailored approaches (*prob.* = 0.13) outperformed advice suggestions which had either very high (*prob.* = 0.052) or low (*prob.* = 0.032) behavioral costs. As most users were located around the middle of the Rasch scale in terms of their energy-saving attitude, tailored advice proved to be only slightly better than a non-personalized guess at the center of the behavioral cost scale (*prob.* = 0.12). Although these results suggest that attitude-tailored advice was more suitable than a non-personalized approach, recommending a measure scaled around the average behavioral costs (δ = 0) would be a good approximation if a user's energy-saving attitude would be unknown.

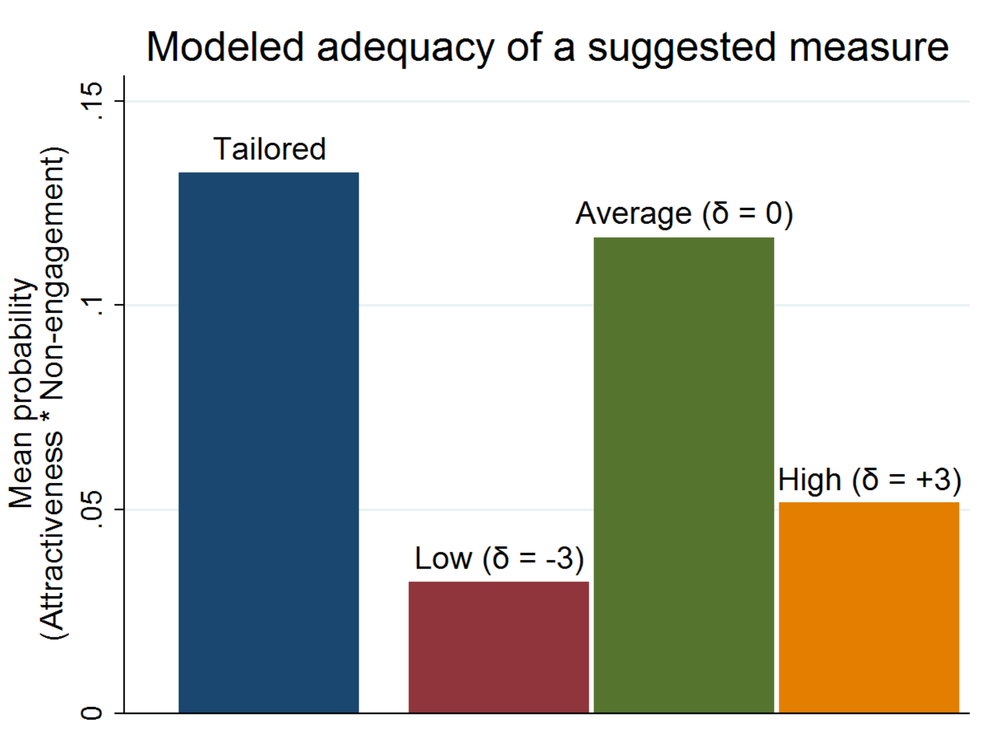

Fig. 5. Modeled adequacy for users who were recommended a single measure. It compares attitude-tailored measures (in blue) with non-tailored recommendations (in red, green, and orange).

4.3. Conclusion

Our second study applied the Rasch scale to an advice context, using a web-based recommender system experiment. We presented three analyses. First, we reliably calibrated a second Rasch scale of energy-saving measures, further confirming the existence of a one-dimensional construct for energy-saving behavior. Second, our rank-ordered logistic regression models pointed out that, when evaluating a list of non-performed measures, our users tend to prefer measures with lower behavioral costs on our scale, in line with the theoretical predictions of Campbell's Paradigm and the Rasch model [22].

Third, we also analyzed the adequacy of four single-measure advice scenarios, comparing attitude-tailored advice to three non-personalized approaches. We posited that whether such advice is adequate or not depends on the trade-off between a measure's feasibility and the probability that it is already performed, while, in turn, these factors are inversely dependent on the attitude-cost difference. In effect, combining these two factors leads us to conclude that conservation advice close to a user's energy-saving attitude is more likely to be adequate than measures that strongly deviate from that attitude, which adheres to a vast body of research that points out the merits of tailored conservation interventions [4]. The employed energy recommender system poses a useful contribution to this line of research, because it is one of the few studies we know of that allows estimation of the comparative effectiveness of tailored and non-tailored interventions.

5. Discussion & policy implications

For many years, both scholars and policy makers have used the curtailment-efficiency dichotomy as a starting point for their respective conceptual research and conservation promotion. However, not only has evidence surfaced that effective energy policy should involve tailored interventions [18], scholars have also suggested that energy-saving behavior should be described as a one-dimensional Rasch construct rather than using two dimensions [23]. The current research has combined these insights in two studies, which have both conceptual and policy implications.

Our first study has yielded a reliable, one-dimensional Rasch scale, consisting of 79 energy-saving measures. This has not only corroborated previous findings that both curtailment and efficiency measures are part of a single dimension [22], we have also done so using a large number of energy-saving options, allowing the use of a Rasch scale in an advice context. Moreover, being able to choose from such a heterogeneous set of measures is paramount for effective conservation advice, as it can appeal to energy-saving consumers of all attitudinal levels, experts and novices alike. In particular, the capacity of the Rasch scale to consider individual differences surpasses the usefulness of a two-dimensional approach, which discerns between behavior types rather than behavioral difficulties.

Our second study, the user experiment, has compared the adequacy of four different types of advice by examining the trade-off between an energy-saving measure's feasibility and redundancy. With regard to the probability of presenting advice that is both new and attractive, attitude-tailored advice outperforms three non-personalized approaches, which suggested a measure from either the low-end, middle or high-end of the behavioral cost scale (see Figure 5). The probability that a low-end suggestion is adequate for a user is found to be particularly small, despite users perceiving measures with low behavioral costs as the most attractive. We reckon that merely recommending such measures is an ill-fated idea from a Rasch model point of view, as they are likely to be already performed. Hence, policy makers who motivate consumers to take short-term curtailment measures, the so-called 'low-hanging fruit' [54], should be cautious of doing so, as their advice may be redundant and possibly annoy the general public. In a similar vein, only suggesting measures with high behavioral costs is found to be only slightly more adequate than a low-cost approach, and is still outperformed by tailored and 'middle-of-the-scale' approaches. It is notable that these high-end measures are found to be mostly state-of-the-art efficiency measures, such as 'install solar PV', which are typically promoted by policy makers to persuade individuals to save energy [cf. 2]. Even though our Rasch model shows that there is high probability that the general public still has to adopt such measures, their low attractiveness reduces their probability to be adequate in our study. Therefore, we suggest policy makers to not only on focus on efficiency investments, but to promote a personalized mix of curtailment and efficiency.

Furthermore, we find a small difference between attitude-tailored advice and an approach that suggests a measure from the middle of the behavioral cost scale (*prob.* = .13 vs *prob.* = .12). Although one could argue this might justify to not personalize advice but to suggest a mid-scale measure as 'install LED lighting' instead (see table 1), thus forgoing on estimating a user's energy-saving attitude, we wish to emphasize the benefits of tailored strategies in preceding research [4], as well as that our attitude estimation procedure is short, only requiring a couple of minutes to make a reliable estimate. Moreover, a middle-scale advice strategy would merely benefit users with an average energy-saving attitude, and not those with a weaker or stronger attitude. However, we do acknowledge that implementing a web-based recommender system that tailors advice to each individual might not be feasible, which would call for alternative, non-personalized strategies. Hence, alternatively, policy makers could consider to divide the Rasch scale in different demographic cohorts and recommend measures that appeal to a particular cohort.

Our research design faces a few limitations. In particular, the rank-order task in the second study may have been too complicated for some users, since 28% of our sample population seems to have misinterpreted the task or has found it difficult to rank nine energy-saving measures. This is further supported by the results of our rank-ordered regression analysis, which revealed a significant presentation order effect. Although future research should attempt to resolve these issues, the employed list length was important for our current design, as it ensured that most users were presented a few measures they did not already perform. Moreover, we have shown that these biases did not have a significant impact on our results, nor did they compromise our main findings on the adequacy of attitude-tailored advice.

We further note that the participant samples used in both studies could be larger and more representative. Although our one-dimensional constructs show good to high item reliability, larger sample sizes, also in terms of the number of items per participant, could increase model reliability. Moreover, participants in Study 1 have been recruited through social media, which might have introduced sample representativeness problems due to self-selection. However, since both there is a strong correlation between constructs, we expect that the findings in this paper can still be generalized to inhabitants of countries similar to the Netherlands.

For future studies on web-based energy platforms, we suggest to draw upon a more diverse set of user evaluation metrics than just a user's attractiveness perception. Considering how a user evaluates a list of suggested measures and the entire decision-support platform have become more common in the field of recommender systems [55]. For instance, studies employing an energy recommender system have shown that a user's system satisfaction can have a positive effect on the energy savings of selected measures [30], as well as that an effective web interface or platform can contribute to the adoption of more energy-efficient behaviors [46]. Furthermore, we encourage the use of social science theories in energy research [56], such as Campbell's Paradigm, to develop effective, personalized conservation approaches. An interesting direction for research would be to test to what extent traditional intervention practices, such as the use of social norms and goal-setting [4,57], are still effective in a personalized recommender platform as the one used in Study 2.

Based on our findings, we recommend policy makers to design energy-saving initiatives which take the capabilities and preferences of consumers into account. The Rasch model has proven to be a valuable tool in predicting which energy-saving measures will be adequate for prospective energy-saving consumers, as the estimated Rasch attitude encompasses these capabilities and preferences. Instead of merely distributing information about energy conservation on websites or through mass-media campaigns [cf. 16], governments should set out interactive tools similar to the one in the present research. The Rasch model could definitely play an important role here, since individually-tailored initiatives are far more promising than any one-size-fits-all approach and estimating an individual's Rasch-based attitude can be done with a simple and short 13-item questionnaire.

Acknowledgements

This work is part of the Research Talent program with project number 406-14-088, which is financed by the Netherlands Organization for Scientific Research (NWO).

References

[1] T. Dietz, G.T. Gardner, J. Gilligan, P.C. Stern, M.P. Vandenbergh, Household actions can provide a behavioral wedge to rapidly reduce US carbon emissions, Proc. Natl. Acad. Sci. 106 (2009) 18452–18456. doi:10.1073/pnas.0908738106.

[2] G.T. Gardner, P.C. Stern, The short list: The most effective actions US households can take to curb climate change, Environ. Sci. Policy Sustain. Dev. 50 (2008) 12–25. doi:10.3200/ENVT.50.5.12-25.

[3] L. Steg, Promoting household energy conservation, Energy Policy. 36 (2008) 4449–4453.

[4] W. Abrahamse, L. Steg, C. Vlek, T. Rothengatter, A review of intervention studies aimed at household energy conservation, J. Environ. Psychol. 25 (2005) 273–291. doi:10.1016/j.jenvp.2005.08.002.

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

[6] G.T. Gardner, P.C. Stern, Environmental problems and human behavior., Allyn & Bacon, 1996. http://psycnet.apa.org/psycinfo/1996-97587-000 (accessed November 9, 2014).

[7] W. Kempton, C.K. Harris, J.G. Keith, J.S. Weihl, Chapter 6: Do Consumers Know "What Works" in Energy Conservation?, Marriage Fam. Rev. 9 (1985) 115–133. doi:10.1300/J002v09n01_07.

[8] P.C. Stern, Psychology and the science of human-environment interactions., Am. Psychol. 55 (2000) 523. doi:10.1037/0003-066X.55.5.523.

[9] S.Z. Attari, M.L. DeKay, C.I. Davidson, W.B. de Bruin, Changing household behaviors to curb climate change: How hard can it be?, Sustain. J. Rec. 4 (2011) 9–11. doi:10.1089/SUS.2010.9724.

[10] S.Z. Attari, M.L. DeKay, C.I. Davidson, W.B. de Bruin, Public perceptions of energy consumption and savings, Proc. Natl. Acad. Sci. 107 (2010) 16054–16059. doi:10.1073/pnas.1001509107.

[11] M. Costanzo, D. Archer, E. Aronson, T. Pettigrew, Energy conservation behavior: The difficult path from information to action, Am. Psychol. 41 (1986) 521–528. doi:10.1037/0003-066X.41.5.521.

[12] P.W. Schultz, Strategies for promoting proenvironmental behavior: Lots of tools but few instructions., Eur. Psychol. 19 (2014) 107–117. doi:http://dx.doi.org/10.1027/1016-9040/a000163.

[13] L. Lutzenhiser, Through the energy efficiency looking glass, Energy Res. Soc. Sci. 1 (2014) 141–151. doi:10.1016/j.erss.2014.03.011.

[14] D. McKenzie-Mohr, Fostering sustainable behavior through community-based social marketing, Am. Psychol. 55 (2000) 531–537. doi:10.1037/0003-066X.55.5.531.

[15] A.H. McMakin, E.L. Malone, R.E. Lundgren, Motivating Residents to Conserve Energy without Financial Incentives, Environ. Behav. 34 (2002) 848–863. doi:10.1177/001391602237252.

[16] C.J. Midden, F.G. Kaiser, L. Teddy McCalley, Technology's four roles in understanding individuals' conservation of natural resources, J. Soc. Issues. 63 (2007) 155–174. doi:10.1111/j.1540-4560.2007.00501.x.

[17] C. Wilson, H. Dowlatabadi, Models of Decision Making and Residential Energy Use, Annu. Rev. Environ. Resour. 32 (2007) 169–203. doi:10.1146/annurev.energy.32.053006.141137.

[18] W. Abrahamse, L. Steg, C. Vlek, T. Rothengatter, The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents, J. Environ. Psychol. 27 (2007) 265–276. doi:10.1016/j.jenvp.2007.08.002.

[19] B. Karlin, N. Davis, A. Sanguinetti, K. Gamble, D. Kirkby, D. Stokols, Dimensions of Conservation Exploring Differences Among Energy Behaviors, Environ. Behav. 46 (2014) 423–452. doi:10.1177/0013916512467532.

[20] B. Gatersleben, L. Steg, C. Vlek, Measurement and Determinants of Environmentally Significant Consumer Behavior, Environ. Behav. 34 (2002) 335–362. doi:10.1177/0013916502034003004.

[21] K. Byrka, Attitude-behavior consistency: Campbell's paradigm in environmental and health domains, (2009). http://www.narcis.nl/publication/RecordID/oai:library.tue.nl:642840 (accessed November 11, 2014).

[22] F.G. Kaiser, K. Byrka, T. Hartig, Reviving Campbell's Paradigm for Attitude Research, Personal. Soc. Psychol. Rev. 14 (2010) 351–367. doi:10.1177/1088868310366452.

[23] J. Urban, M. Ščasný, Structure of Domestic Energy Saving: How Many Dimensions?, Environ. Behav. 48 (2016) 454–481. doi:10.1177/0013916514547081.

[24] T.G. Bond, C.M. Fox, Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edition, Psychology Press, 2006.

[25] S.E. Embretson, S.P. Reise, Item Response Theory for Psychologists, Psychology Press, 2000.

[26] R.M.J. Benders, R. Kok, H.C. Moll, G. Wiersma, K.J. Noorman, New approaches for household energy conservation—In search of personal household energy budgets and energy reduction options, Energy Policy. 34 (2006) 3612–3622. doi:10.1016/j.enpol.2005.08.005.

[27] S. Darby, Energy advice–what is it worth, Proc. Eur. Counc. Energy-Effic. Econ. Summer Study Pap. III. 5 (1999) 3–5.

[28] H.A. He, S. Greenberg, E.M. Huang, One Size Does Not Fit All: Applying the Transtheoretical Model to Energy Feedback Technology Design, in: Proc. SIGCHI Conf. Hum. Factors Comput. Syst., ACM, New York, NY, USA, 2010: pp. 927–936. doi:10.1145/1753326.1753464.

[29] G. Brandon, A. Lewis, Reducing household energy consumption: a qualitative and quantitative field study, J. Environ. Psychol. 19 (1999) 75–85. doi:10.1006/jevp.1998.0105.

[30] B.P. Knijnenburg, M. Willemsen, R. Broeders, Smart sustainability through system satisfaction: Tailored preference elicitation for energy-saving recommenders, in: Proc. 20th Am. Conf. Inf. Syst. AMCIS 2014 Smart Sustain. Inf. Syst. Oppor., Savannah, Georgia, United States, 2014. http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1138&context=amcis2014 (accessed October 28, 2016).

[31] F.G. Kaiser, M. Wilson, Goal-directed conservation behavior: the specific composition of a general performance, Personal. Individ. Differ. 36 (2004) 1531–1544. doi:10.1016/j.paid.2003.06.003.

[32] I. Ajzen, The theory of planned behavior, Organ. Behav. Hum. Decis. Process. 50 (1991) 179–211. doi:10.1016/0749-5978(91)90020-T.

[33] D.T. Campbell, Social Attitudes and Other Acquired Behavioral Dispositions, in: Psychol. Study Sci. Study II Empir. Substruct. Relat. Sci. Vol. 6 Investig. Man Socius Their Place Psychol. Soc. Sci., McGraw-Hill, New York, NY, USA, 1963: pp. 94–172. http://dx.doi.org/10.1037/10590-003.

[34] A.H. Eagly, S. Chaiken, The psychology of attitudes., Harcourt Brace Jovanovich College Publishers, 1993. http://psycnet.apa.org/psycinfo/1992-98849-000 (accessed April 15, 2014).

[35] C. Dienes, Actions and intentions to pay for climate change mitigation: Environmental concern and the role of economic factors, Ecol. Econ. 109 (2015) 122–129. doi:10.1016/j.ecolecon.2014.11.012.

[36] W. Greve, Traps and gaps in action explanation: Theoretical problems of a psychology of human action., Psychol. Rev. 108 (2001) 435.

[37] K.C.H.J. Smolders, Y.A.W. de Kort, A.D. Tenner, F.G. Kaiser, Need for recovery in offices: Behavior-based assessment, J. Environ. Psychol. 32 (2012) 126–134. doi:10.1016/j.jenvp.2011.12.003.

[38] T.M.M. Verhallen, R.G.M. Pieters, Attitude theory and behavioral costs, J. Econ. Psychol. 5 (1984) 223–249. doi:10.1016/0167-4870(84)90024-2.

[39] G. Rasch, On General Laws and the Meaning of Measurement in Psychology, in: The Regents of the University of California, 1961. http://projecteuclid.org/euclid.bsmsp/1200512895 (accessed October 17, 2014).

[40] M. Reckase, Multidimensional item response theory, Springer, 2009. http://link.springer.com/content/pdf/10.1007/978-0-387-89976-3.pdf (accessed November 2, 2016).

[41] F.G. Kaiser, B. Oerke, F.X. Bogner, Behavior-based environmental attitude: Development of an instrument for adolescents, J. Environ. Psychol. 27 (2007) 242–251. doi:10.1016/j.jenvp.2007.06.004.

[42] J.A. Konstan, J. Riedl, Recommended for you, IEEE Spectr. 49 (2012) 54–61. doi:10.1109/MSPEC.2012.6309257.

[43] L. Brozovsky, V. Petricek, Recommender system for online dating service, ArXiv Prepr. Cs0703042. (2007). http://arxiv.org/abs/cs/0703042 (accessed December 12, 2016).

[44] F. Ricci, L. Rokach, B. Shapira, Introduction to recommender systems handbook, Springer, 2011. http://link.springer.com.dianus.libr.tue.nl/chapter/10.1007/978-0-387-85820-3_1 (accessed November 11, 2014).

[45] C. Fischer, Feedback on household electricity consumption: a tool for saving energy?, Energy Effic. 1 (2008) 79–104. doi:10.1007/s12053-008-9009-7.

[46] A. Starke, M. Willemsen, C. Snijders, Effective User Interface Designs to Increase Energy-efficient Behavior in a Rasch-based Energy Recommender System, in: Proc. Elev. ACM Conf. Recomm. Syst., ACM, New York, NY, USA, 2017: pp. 65–73. doi:10.1145/3109859.3109902.

[47] B.P. Knijnenburg, M.C. Willemsen, Understanding the effect of adaptive preference elicitation methods on user satisfaction of a recommender system, in: Proc. Third ACM Conf. Recomm. Syst., 2009: pp. 381–384. http://dl.acm.org.dianus.libr.tue.nl/citation.cfm?id=1639793 (accessed January 4, 2014).

[48] K.E. Green, C.G. Frantom, others, Survey development and validation with the Rasch model, in: Int. Conf. Quest. Dev. Eval. Test. Charlest. SC, 2002.

[49] J.M. Linacre, Winsteps® Rasch measurement computer program, Winsteps.com, Beaverton, Oregon, 2016. http://www.winsteps.com (accessed January 1, 2016).

[50] J.M. Linacre, Optimizing rating scale category effectiveness, J Appl Meas. 3 (2002) 85–106.

[51] E.V. Smith Jr., Understanding Rasch measurement: Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals, J. Appl. Meas. 3 (2002) 205–231.

[52] S. Beggs, S. Cardell, J. Hausman, Assessing the potential demand for electric cars, J. Econom. 17 (1981) 1–19. doi:10.1016/0304-4076(81)90056-7.

[53] G.N. Punj, R. Staelin, The Choice Process for Graduate Business Schools, J. Mark. Res. 15 (1978) 588–598. doi:10.2307/3150628.

[54] M.P. Vandenbergh, J. Barkenbus, J.M. Gilligan, Individual Carbon Emissions: The Low-Hanging Fruit, Social Science Research Network, Rochester, NY, 2008. https://papers.ssrn.com/abstract=1161143 (accessed December 22, 2016).

[55] B.P. Knijnenburg, M.C. Willemsen, Z. Gantner, H. Soncu, C. Newell, Explaining the user experience of recommender systems, User Model. User-Adapt. Interact. 22 (2012) 441–504. doi:10.1007/s11257-011-9118-4.

[56] B.K. Sovacool, S.E. Ryan, P.C. Stern, K. Janda, G. Rochlin, D. Spreng, M.J. Pasqualetti, H. Wilhite, L. Lutzenhiser, Integrating social science in energy research, Energy Res. Soc. Sci. 6 (2015) 95–99. doi:10.1016/j.erss.2014.12.005.

[57] O. Iweka, S. Liu, A. Shukla, D. Yan, Energy and behaviour at home: A review of intervention methods and practices, Energy Res. Soc. Sci. 57 (2019) 101238. doi:10.1016/j.erss.2019.101238.