

The Interplay between Food Knowledge, Nudges, and Preference Elicitation Methods Determines the Evaluation of a Recipe Recommender System

Ayoub El Majjodi¹, Alain D. Starke^{2,1}, Mehdi Elahi¹ and Christoph Trattner¹

¹MediaFutures, University of Bergen, 5007 Bergen, Norway

²ASCoR, University of Amsterdam, Nieuwe Achtergracht 166, 1001 NG Amsterdam, Netherlands

Abstract

Domain knowledge can affect how a user evaluates different aspects of a recommender system. Recipe recommendations might be difficult to understand, as some health aspects are implicit. The appropriateness of a recommender's preference elicitation (PE) method, whether users rate individual items or item attributes, may depend on the user's knowledge level. We present an online recipe recommender experiment. Users ($N = 360$) with varying levels of subjective food knowledge faced different cognitive digital nudges (i.e., food labels) and PE methods. In a 3 (recipes annotated with no labels, Multiple Traffic Light (MTL) labels, or full nutrition labels) x 2 (PE method: content-based PE or knowledge-based) between-subjects design. We observed a main effect of knowledge-based PE on the healthiness of chosen recipes, while MTL label only helped marginally. A Structural Equation Model analysis revealed that the interplay between user knowledge and the PE method reduced the perceived effort of using the system and in turn, affected choice difficulty and satisfaction. Moreover, the evaluation of health labels depends on a user's level of food knowledge. Our findings emphasize the importance of user characteristics in the evaluation of food recommenders and the merit of interface and interaction aspects.

Keywords

Recommender Systems, Food, Digital nudges, Nutrition labels, Preference Elicitation

1. Introduction

An increasing number of food decisions are made digitally [1]. In addition to online grocery stores, recipe websites play a pivotal role in supporting home cooking [2]. At the same time, users of such websites may find it difficult to navigate a large number of recipes. Some recipes are more challenging to understand or cook than others [3], where users may lack sufficient food knowledge or skills to engage with them [4]. For example, while some users have clear preferences regarding specific recipe features, such as cooking time and the number of ingredients, others may make food choices based on past positive experiences with a recipe and seek out 'more like this' [5].

The large availability of recipes requires the use of information-filtering systems. Food recommender systems have played an instrumental role in the popularity of recipe website [6],

IntRS '23: 10th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems at the 17th ACM Conference on Recommender Systems (RecSys '23), September 18–22, 2023, Singapore

✉ ayoub.majjodi@uib.no (A. E. Majjodi); a.d.starke@uva.nl (A. D. Starke); mehdi.elahi@uib.no (M. Elahi); christoph.trattner@uib.no (C. Trattner)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

largely focusing on predictive accuracy [7, 2]. However, eating is a particularly challenging domain to achieve robust improvements in accuracy when aiming to go beyond popularity-based approaches. Food preferences strongly depend on various factors, such as context [8], group size [9], time of the day, and the day of the week [2].

Improvements may therefore be sought in other recommender aspects. User preferences are also formed by how information is presented, combined with the user's level of understanding [10, 9]. While users may have a specific eating goal (e.g., health or sustainability [5]), the information presented by recommenders may be too limited to make an informed decision [11]. To this end, summary labels, as also used on supermarket products, may help to grasp the healthiness or nutritional content of a recipe, particularly for novice users.

The user's knowledge may also affect how a recommender's preference elicitation method is evaluated. In the energy conservation domain, a domain subject to behavioral change [12], a user's evaluation of a preference elicitation (PE) method seems to depend on the user's level of domain knowledge [10, 13]. Users that do not know much about energy conservation tend to be more satisfied when using a case-based PE method, in which individual items are (dis)liked. In contrast, experienced users are found to be more satisfied when interacting with attributes of energy-saving measures, such as effort and investment costs.

We argue that a similar dynamic applies to the food recommender domain. We expect novice users of recipe recommenders and websites to incrementally explore recipes based on what they liked previously, as one would in a content-based method [7]. In contrast, experienced users would seek out new recipes more easily based on preferred attributes, as one would through a knowledge-based method [14].

This paper examines the role of user knowledge on decisions in a recipe recommender system. Instead of following an algorithmic optimization approach that only focuses on user preferences [7], we examine the influence of two other recommender aspects: food labels and preference elicitation methods. A user's domain knowledge is critical here, regarding the most optimal interface representation or interaction method.

First, we apply digital, informational nudges to recommendation lists in the form of nutrition and health labels. In the food domain, these have been used primarily in supermarkets to communicate nutrition information to consumers in a simplified manner [15, 16]. Most notable are summary labels, such as the Nutriscore to classify foods between A and E. Most popular is the Multiple Traffic Light (MTL) label [17], which uses colors to indicate (un)healthy intakes of four nutrients: fat, saturated fat, sugar, and salt. This study considers two labels, with varying degrees of difficulty: The MTL label and the back-of-pack nutritional facts [17].

Second, we examine the role of two different PE methods. First, we develop a content-based recommender in which a user picks a favorite recipe from a randomly generated list of recipes. Second, we develop a knowledge-based recommender system in which users indicate their preferences for recipes based on a set of attributes, such as cooking time and difficulty. We expect that user choices might be affected by the use of labels, and the evaluation depends on the interplay between user knowledge levels and either the labels or PE methods used. We formulate the following research questions:

- **RQ1:** To what extent do different nutrition labels support healthier recipe choices?

- **RQ2:** Does the user evaluation of a recipe recommender system depend on the interplay between food knowledge and different preference elicitation methods?

2. Related Work

2.1. Food Recommender Systems

The field of recommender systems has received considerable research attention due to the complex and fundamental nature of food [6, 2]. However, most studies have focused on improving prediction accuracy [9]. To this end, various techniques have been explored. While content-based approaches are initially found to outperform other methods (e.g., collaborative, knowledge-based) [7], collaborative filtering has also yielded better results in other studies [2].

A recurring problem is to support healthier food and recipe choices. There is an apparent tradeoff between ‘user preferences’ and health in recipe recommendation [6], particularly due to the popularity of unhealthy recipes [1]. Rather than restricting content based on health, various studies have examined hybrid solutions (e.g., through post ranking on health [18]) and interface solutions. For example, multi-list interfaces have been developed to support healthier eating goals, where multi-list recommenders are evaluated more favorably than single-list interfaces [5].

Nonetheless, very few studies have explored the impact of preference elicitation methods on user evaluation, particularly not in relation to a user’s knowledge level. Research that presents novel preference elicitation (PE) methods typically involve conversational recipe recommenders [19]. For example, two studies have compared the effect of using different modalities on user choice or evaluation [20, 21]. While these modalities might affect how a user experiences an interaction or even what item is chosen, these modality types seem to bear no relation to a user’s knowledge level. Thus, the current study aims to fill this gap by empirically examining the effects of different preference elicitation methods and subjective food knowledge on the user experience.

2.2. Adaptive Preference Elicitation Methods

Across all recommender domains, several preference elicitation (PE) approaches have been proposed and evaluated [2]. One of the earlier works in the food domain was [7], following a rather simplistic procedure where recipe ratings are obtained from users and transferred into ingredient ratings. While the user could still interact with individual recipes (as also in [22]), the system then aggregated the ratings of the ingredients to generate rating predictions.

More extensive approaches would include user preferences for individual recipes and attributes. Elahi et al. [23] consider additional factors when building a recommendation model, by including user food preferences, nutritional indicators, and ingredient costs. This results in a model that combines the predicted value of a recipe along with the above-noted factors to generate recommendations.

Food recommender studies do not explicitly discern between PE methods. In the end, most methods are simply a requirement for following a specific recommender model. Knowledge-based food recommender studies elicit extensive recipe attribute preferences [14, 24], while

content-based and collaborative approaches can deal with interactions at the individual recipe level. To date, besides a preliminary study [22], it has not been considered that specific users, based on their knowledge or capabilities, may prefer specific PE methods or require specific information to be presented in a food recommender interface.

The interplay between user characteristics and PE methods is investigated in the energy conservation domain. Knijnenburg and Willemsen [10] compare two preference elicitation methods for a household energy recommender system. Users of a case-based PE method could indicate to (dis)like individual energy-saving measures (e.g., ‘turn off the lights after leaving a room’). In contrast, users of an attribute-based PE method could either decrease or increase the weights of different energy-saving attributes, such as ‘investment costs’ or ‘effort’. They find, also in follow-up studies [13], that users with high domain knowledge are more satisfied when using an attribute-based PE method, while users with lower domain knowledge prefer case-based PE. As food PE methods can also be differentiated in terms of interacting with individual recipes (i.e., case-based, such as in content-based recommendation) and recipe features (i.e., feature-based, such as in knowledge-based recommendation), we expect to observe an interaction effect between food knowledge and the PE method on how a recommender is evaluated.

2.3. Nutrition Labels and Digital Nudging

Nutritional information about food and recipes might not always be apparent to users. This is another area where the user’s domain knowledge may affect their preferences, particularly in cases where it is either emphasized or not.

One way to communicate the healthiness of foods and recipes is through nutrition labels [17]. One category is back-of-package labels that outline the nutritional contents of a food product in detail [25]. Another category is front-of-package (FoP) labels that are used for both products and recipes, which typically summarize the product’s or recipe’s nutritional content. For example, this could be by highlighting specific nutrients or aggregating nutrition information towards a score [17, 25].

The health benefits of such FoP labels have been shown longitudinally, supporting healthy food intake [26]. While a growing body of evidence supports the effectiveness of FoP nutrition labels in promoting healthy food choices in physical settings, the impact of nutrition labels in digital contexts has been relatively understudied [16]. In particular, little is known about the effectiveness of FoP labels in personalized environments [27].

FoP labels can be regarded as a digital nudge, a change in an interface that leads to predictable choices [28]. More specifically, such a label is a cognitively oriented healthy eating nudge [15], as users are encouraged to re-consider their preferences and choices based on deliberation. Some of these labels, such as the Multiple Traffic Light (MTL) label, is also accompanied by a coloring system, which supports intuitive decision-making [29].

Various digital nudges mainly relate to interface aspects of food recommenders. For example, visual and textual explanations have been shown to shift user preferences towards healthier recipes [30, 24]. This study mainly builds upon earlier work where ‘boosting’ is examined to first explain FoP labels to users, after which recipes are annotated with them [22]. This has led to a higher proportion of healthier choices in the recommender interface. Instead of applying

'boosts' in this study, we seek to examine the effectiveness of two cognitive, informational nudges. To examine the differentiating effect of user knowledge when combined with nudges, we either annotate recipes with back-of-package labels (i.e., Nutritional Facts label¹) or Multiple Traffic Light (MTL) labels [31]. We expect that users with higher levels of domain knowledge are better able to understand the full back-of-package label, compared to the MTL label also being appropriate for low-knowledge users.

2.4. Objectives

We extend previous work of food recommender systems [6], by examining the role of preference elicitation methods and digital nudges (through nutrition labels) [28, 10, 13]. First, we investigate how different labeling systems can facilitate healthier decision-making when selecting recipes (RQ1). We compare two label-based scenarios (with either an MTL label or a 'full' back-of-pack label) with a no-label baseline, focusing on the interplay between labels and the preference elicitation method and how this affects the user's evaluation. In doing so, we also consider a user's knowledge level and the preference elicitation method.

Second, we examine the impact of the interplay of user knowledge and preference elicitation methods on user choice and evaluation. We differentiate between two methods, content-based and knowledge-based, being 'case-based' or 'attribute-based' PE methods, respectively; in line with Knijnenburg et al. [13]. For the content-based approach, users are asked to indicate whether they like individual recipes, while the knowledge-based approach elicits preferences based on recipe features and personal characteristics, such as cooking time and self-reported weight goals. In line with Knijnenburg and Willemsen [32], we used Structural Equations Modeling (SEM) to construct a path model, in which changes to the recommender were related to perception aspects and, in turn, experience aspects.

3. Study Design

3.1. Dataset

To address our research questions, we used a dataset from the popular recipe website All-recipes.com. From the larger corpus of 58,000 recipes, we sampled 5,000 recipes from different food categories for the main dish. In addition to the recipe title, all nutrients required to build the recommender were extracted.

3.2. Recommender Approaches

We employed two recommendation approaches, both of which rely on explicit preference elicitation methods [9].

3.2.1. Content-based (CB)

The content-based recommender system generated recommendations based on similarity with recipes liked by the user. It employed the Term Frequency-Inverse Document Frequency (TF-

¹<https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label>

IDF) model to generate personalized recommendations based on the recipe’s ingredients. The ingredient list was vectorized, operationalizing TF as the weight (per 100g) present in the recipe. To build a user model in our study, we presented a list of 10 recipes to the user that included detailed descriptions of their ingredients, pictures, servings, and calorie information. For computing the final recommendations for the user, we computed similarities between the user and item profile, employing a cosine similarity metric and the ingredient vectors (cf. [14]). This method adhered to the standard methods in food recommender systems and had been shown to generate decent results in the domain of food recommendations where no collaborative filtering is possible [33].

3.2.2. Knowledge-based (KB)

We developed a knowledge-aware recommender system that extended the work of Musto et al. [14]. An overview of elicited features is described in Table 1. Users were asked to disclose personal characteristics and practical and health-related preferences related to recipes. A score-based ranker used encoded knowledge relations between user factors and recipe features to recipes, based on the user’s profile. The scores are adjusted based on metadata such as ingredients or nutritional value. Table 1 also presents the rules to score recipes based on the user needs, which is among others based on information found in [34, 35, 36, 24].

Table 1

User Factors and rules to score and generate relevant recipes. The recipes were subject to three groups (higher, normal, and lower value) based on the average values of nutritional and practical aspects.

User factor	Ranges	Recipe attributes affected
BMI	< 18	higher calories, higher protein, higher carbohydrates
	18-24	normal calories
	> 24	lower calories, lower fat
Eating Goals	lose weight	higher calories, higher protein, higher carbohydrate
	neutral	normal calories
	gain weight	lower calories, lower fat
Behaviours	> 30min/day activity	higher protein, higher calories
Sleep	< 7h hours	higher magnesium, higher vitamin B6, lower fat, higher protein
Depressed feelings	yes/no	higher protein, lower carbohydrates
Cooking experience	high, medium, low	number of instructions, number of ingredients
Cooking time	30 min, 40-60 min, > 60 min	preparation time, number of ingredients

3.3. Research Design and System Procedure

The participants were assigned to a between-subjects design with a 2 (Preference Elicitation (PE): Content-Based (CB) vs. Knowledge-Based (KB)) x 3 (labeling systems: no label vs. Multiple Traffic Light (MTL) vs. Full label) configuration. In the content-based method, users selected their preferred recipe from randomly generated options, while the knowledge-based condition involved users providing health and food-related information. Personalized recipes were then labeled with: No label, MTL label, or Full labels, as depicted in Figure 1 (A-C).

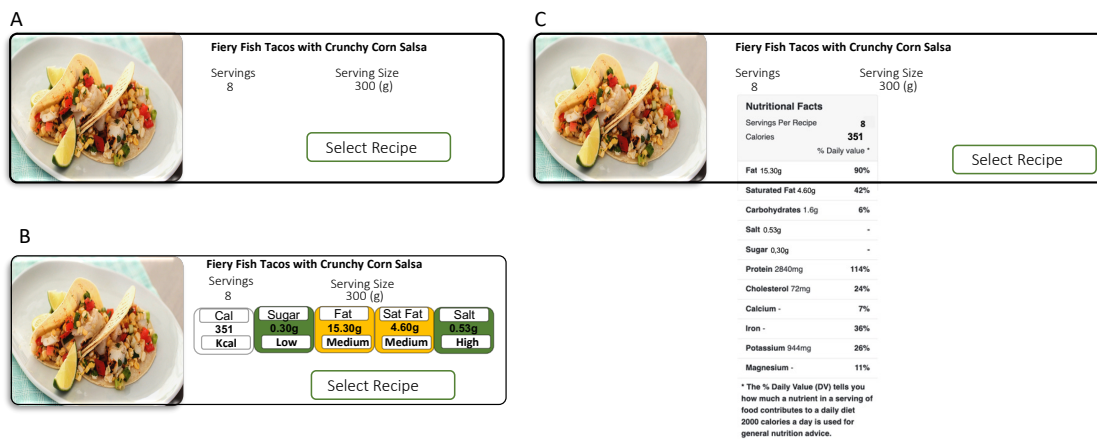


Figure 1: An example of personalized recipes annotated with No label (A), Multiple Traffic Light label (B), or Full label (C).

We implemented a user flow that started with obtaining consent from the study participants. The whole online user study flow is illustrated in Figure 2. Once the participants agreed to take part in the study, they provided basic demographic information, including age group, education level, and gender. Information processing was in line with Ethical guidelines at University of Bergen, Norway. In both the content and knowledge-based conditions, the choice task and evaluation questionnaire were similar, with users choosing a single preferred recipe from the list of recommended items (top-10 list). In both conditions, recipes were labeled either with an MTL label, a Full label or no label. Finally, the participants evaluated the system based on the performed choice regarding satisfaction, difficulty, and effort.

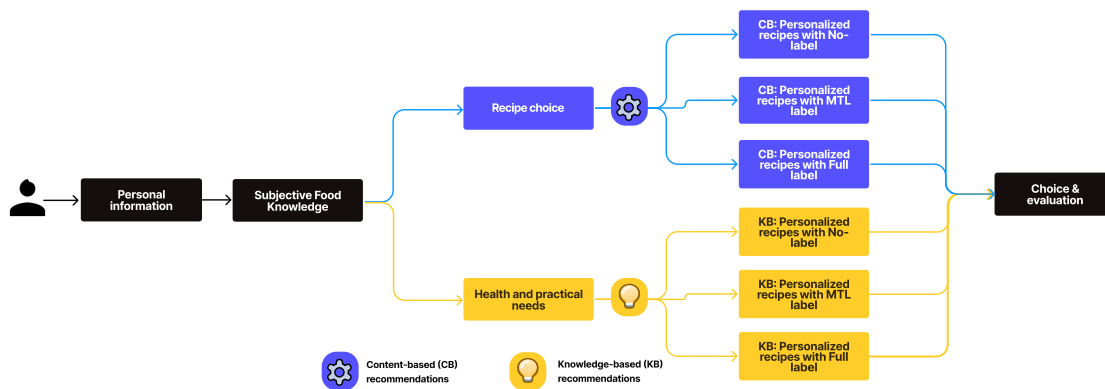


Figure 2: The User flow for the online experiment.

3.4. Participants

We utilized the Prolific crowdsourcing platform to recruit users for our study, offering 0.80 GBP as compensation. A total of 360 participants took part in the study. However, after pre-screening

the data, we had to exclude 54 participants. Grounds for exclusion were based on multiple violations of non-attentiveness: not disclosing realistic knowledge-based criteria (e.g., a weight of 15kg), providing uniform responses in the user evaluation questionnaires, and/or completing the study in under 2 minutes. Our final analysis was performed on a sample of 306 (65% female) participants split equally on study conditions, with an average age of 30.5 years.

3.4.1. Ethical Statement

This research adhered to the ethical guidelines of the University of Bergen and the Norwegian guidelines for scientific research. It was judged to pass without further extensive review, for it contained no misleading information, stress tasks, nor would it elicit extreme emotions.

3.5. Measures

3.5.1. Recipe Healthiness

To determine the healthiness of the recipes in our dataset, we utilized the FSA score. It was introduced by the UK Food Standards Agency [31] and is considered a reliable measure to estimate recipes' healthiness. It was successfully used in multiple human-computer interaction and recommender systems studies on recipes [1, 22, 20, 18]. The metric represented an inverse healthiness score and ranged from 4 (very healthy) to 12 (not healthy). It was based on the levels of fat, saturated fat, sugar, and salt per 100g in a recipe, adhering to nutritional intake guidelines. For our algorithmic sampling, we classified recipes as healthy up to a score of 8, while higher scores were designated as unhealthy.

3.5.2. Food knowledge and user evaluation

To measure the users' nutritional knowledge levels, we employed the Subjective Food Knowledge (SFD) questionnaire, which was validated in prior studies [37, 38]. The SFD questionnaire comprised five items, that are rated on a five-point Likert scale.

A user's experience of using our system was assessed through the recommender system evaluation framework Knijnenburg and Willemsen [32]. For this study, we expected changes in terms of how effortful a user perceived the interaction to be, while also inquiring on two different choice outcomes: choice difficulty and choice satisfaction. These two experience aspects are commonly used to evaluate recommender interactions [32, 13]. All questionnaire items used for the user evaluation were previously validated in relevant domains through earlier studies: for perceived effort [39], choice difficulty [40, 5], and choice satisfaction [12, 5].

All items were submitted to a confirmatory factor analysis. Subjective food knowledge was analyzed separately to allow for interaction effects, while the other aspects were inferred as part of a structural equation model analysis. Table 2 describes the factor loadings and Cronbach's Alpha, showing that items with low loadings (indicated in grey) were excluded from the analysis. All aspects adhered to internal consistency guidelines ($\alpha > .70$), while they also met guidelines for convergent validity based on the average variance explained ($AVE > 0.5$).

Table 2

Results of the confirmatory factor analysis across different user characteristics and experience aspects. Items were measured on 5-point Likert scales. To allow for interaction effects in the path model, subjective food knowledge was analyzed separately. Cronbach's Alpha is denoted by α , items in gray were omitted from analysis, due to low factor loading .

Aspect	Item	Loading
Subj. Food Knowledge <i>AVE</i> = .744 α = .872	Compared with an average person, I know a lot about healthy eating.	.836
	I think, I know enough about healthy eating to feel pretty confident when choosing a recipe.	.879
	I know a lot about food to evaluate the healthiness of a recipe.	.847
	I do not feel very knowledgeable about healthy eating.	-.887
Choice Satisfaction <i>AVE</i> = .616 α = .795	I like the recipe I have chosen.	.750
	I think I will prepare the recipe I have chosen.	.637
	The chosen recipe fits my preference.	.746
	I know many recipes that I like more than the one I have chosen.	
Choice Difficulty <i>AVE</i> = .554 α = .659	I would recommend the chosen recipe to others.	.681
	I changed my mind several times before making a decision.	
	Making a choice was overwhelming.	.766
Perceived Effort <i>AVE</i> = .504 α = .632	It was easy to make this choice.	-.549
	The system takes up a lot of time.	.851
	I quickly understood the functionalities of the system.	-.683
	Many actions were required to use the system.	.557

4. Results

We examined our research questions through two different analyses. First, we examined the healthiness of user choices (RQ1) through a two-way ANCOVA, predicting the FSA score based on our research design. Second (RQ2), we investigated how users evaluated different labels and preference elicitation (PE) methods through Structural Equation Modelling, also assessing mediated relations. This analysis was performed in line with the recommender system user experience framework [32], relating our research design (objective system aspects) to user perception (effort) and experience (choice difficulty and satisfaction) while considering user characteristics (food knowledge) and behavior (healthiness of recipes chosen).

4.1. RQ1: Healthiness of Chosen Recipes

We predicted the FSA score of chosen recipes based on the labels presented and the user preference elicitation method. We also included food knowledge as a covariate and examined possible interaction effects with PE or labels, but did not observe any. Descriptive statistics indicated 64% of recipes were chosen in the MTL condition, compared to the 54% for the full labeling system.

The results of the two-way ANCOVA are presented in Table 3. The FSA score of chosen recipes was found to not significantly depend on the type of nutritional label presented: $F(2, 299) = 2.93$, $p = 0.055$. As the relatively small p -value shows, we did observe small differences across conditions, where the healthiest choices were made when facing MTL labels ($M_{FSA} = 7.17$, $SD_{FSA} = 2.01$), while scores were higher in the baseline ($M_{FSA} = 7.64$, $SD_{FSA} = 2.01$), and the full label condition ($M_{FSA} = 7.70$, $SD_{FSA} = 2.05$). However, these were not significant, suggesting that in a personalized choice context, cognitively oriented labelling nudges could not further support healthier recipe choices.

Table 3

Results of two-way ANCOVA tests, predicting the healthiness of chosen recipes across different labeling conditions and preference elicitation methods. Food knowledge was added as a covariate. *** $p < .001$, ** $p < .01$, * $p < .05$.

Factor (FSA score)	df	F
Model	6	4.14**
Nutrition labels (No Label-MTL-Full label)	2	2.93
Preference Elicitation (CB-KB)	1	8.30**
Labels * Preference Elicitation	2	.75
Food Knowledge	1	8.61**

We did observe that the preference elicitation method employed had a significant effect on the healthiness of choices made. Participants using the knowledge-based method ($M_{FSA} = 7.14$, $SD_{FSA} = 2.03$) made healthier choices than those using a content-based approach ($M_{FSA} = 7.85$, $SD_{FSA} = 1.98$): $F(1, 299) = 8.30$, $p = 0.004$. This suggested that a knowledge-based PE allowed users to find recipes more easily. Our analysis further revealed that there was no significant interaction effect between the type of labels used and the preference elicitation method employed, which is also depicted in Figure 3. Finally, the two-way ANCOVA also revealed a relation between food knowledge and the FSA score, indicating that users with higher levels of food knowledge made healthier choices ($r(306) = -.17$).

4.1.1. Conclusion

We found that annotating personalized recipes with either a Multiple Traffic Light or a Full label did not significantly lead to healthier recipe choices. Although previous studies suggested the possible merit of labels in recommender systems [27, 22], we only observed a small, non-significant ($p = 0.054$) improvement, particularly when using a front-of-pack Multiple Traffic Light label. In contrast, the full, back-of-pack nutrition facts label led to similar outcomes as in the baseline. As no interactions with food knowledge were found, this suggested that such cognitive digital nudges are less effective in a personalized recommender context, and choice outcomes do not depend on a label's understandability.

We did observe that a knowledge-based preference elicitation (PE) method led to healthier recipe choices. This indicated that a knowledge-based PE method could support users to make healthier choices, particularly by generating healthier recommendations as a translation

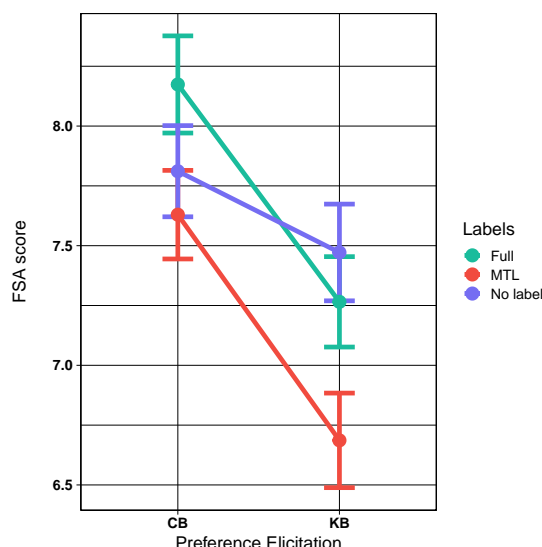


Figure 3: Mean FSA scores (i.e., inverse healthiness scores) of chosen recipes, across different conditions. Error bars depict 1 S.E.

to the user greater control in guiding the recommendations process in this recommender approach. Moreover, users with higher food knowledge made healthier choices. Although we did not observe any interaction effects on choice, we explored this relationship again in the next subsection, examining the interplay between knowledge level, PE methods and labelling conditions on perception and experience aspects.

4.2. RQ2: User Evaluation and Preference Elicitation Methods

To examine how users evaluated different interaction methods and interface nudges based on their knowledge level, we formed a Structural Equation Modeling (SEM). All objective and subjective aspects, along with user characteristics and interaction metrics, were organized in a path model. Following the guidelines by Knijnenburg and Willemsen [32], we first fitted a fully saturated model, organizing the path from objective aspects (i.e., the conditions) to perception (i.e., effort) and experience aspects (i.e., choice difficulty and satisfaction), after which non-significant relations were omitted.

The resulting model is presented in Figure 4, showing a good fit: $\chi^2(94)$, $p < .01$, $CFI = .963$, $TLI = .955$, $RMSEA = .038$, 90%-CI: [.022; .052]. The model met the guidelines for discriminant validity, as the correlations between latent constructs were smaller than the square root of each construct's Average Variance Explained (AVE). Please note that the chosen FSA score was included in our analysis, but it was not related to any mediated path.

Figure 4 depicts two paths towards choice satisfaction, stemming from the objective aspects. First, we observed an interaction effect between the use of an MTL label and a user's subjective food knowledge, in addition to a main effect of MTL. These effects be best explained through the marginal effects plot in Figure 5. We found that people facing MTL labels were on average, less satisfied with the recipe they had chosen than users in our conditions ($coef. = -1.412$,

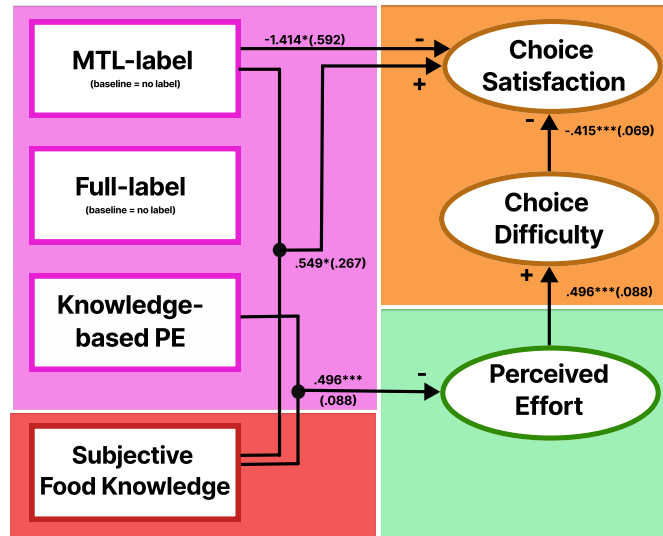


Figure 4: Structural Equation Model (SEM). Numbers on the arrows represent the β -coefficients, and standard errors are denoted between brackets. Effects between the subjective constructs are standardized and can be considered as correlations. Other effects show regression coefficients. Objective system aspects are purple, perception aspects are green, and experience aspects are orange. User characteristics are red. $***p < .001$, $**p < .01$, $*p < .05$.

$p = 0.017$). However, the interaction effect with food knowledge showed that this particularly applied to users with low knowledge levels, as choice satisfaction significantly increased among users with a higher knowledge level facing MTL labels ($coef. = .547$, $p = 0.041$).

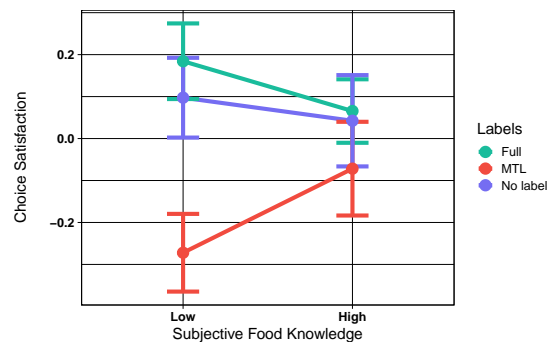


Figure 5: Standardized score for choice satisfaction across different labeling conditions and subjective food knowledge levels. Errors bars represent 1 S.E.

The path towards perceived effort stemmed from the interaction between the PE method and food knowledge. Higher levels of food knowledge led to lower levels of perceived effort among those facing a knowledge-based recommender ($coef. = -.436$, $p < .05$). To better understand this effect, please inspect Figure 6a, which depicts a two-sided interaction effect. For a KB recommender, effort was slightly reduced for users with higher knowledge levels. For a CB recommender, in contrast, perceived effort increased among users with higher knowledge levels.

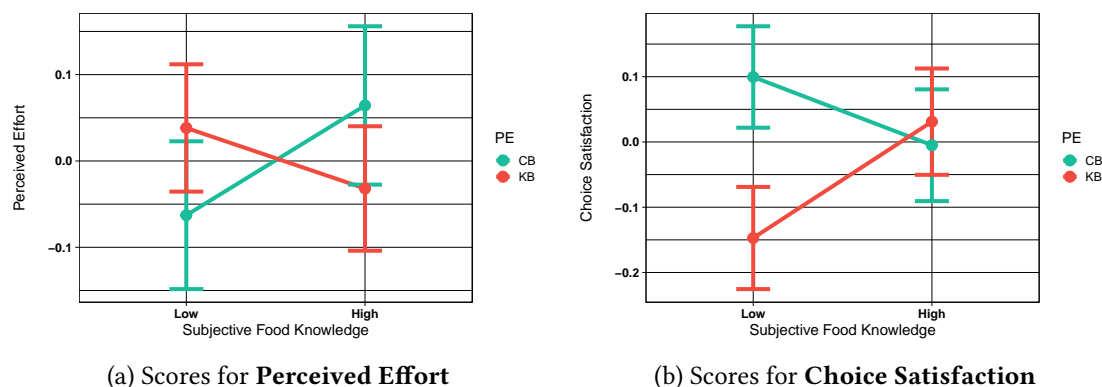


Figure 6: Standardized Scores for Perceived Effort (a) and Choice Satisfaction (b), across different preference elicitation (PE) methods and subjective food knowledge levels. Errors bars represent 1 S.E.

This suggested that the perception of recommender use depended on the interplay between knowledge and the PE method, in line with [13].

The perceived effort, in turn, affected the experience aspects. We observed a positive relationship between effort and choice difficulty ($coef. = .496, p < .001$), suggesting that effortful interactions were also related to harder decision-making processes. The full path towards choice difficulty, as depicted in Figure 4, shows that effort decreased due to the interplay between knowledge-based PE and knowledge, which in turn was related to choice difficulty. However, a test of indirect effects did not reveal significant support for a path towards choice difficulty that was fully mediated by perceived effort: $\beta = -.216, p = .053$, even though a lack of statistical power may have undermined this test.

Finally, we observed a negative relation between choice difficulty and choice satisfaction ($coef. = -.415, p < .001$). This relation was observed in various previous studies (cf. [32]), but here it highlighted that the effects of the interplay between the PE method and user knowledge affected different evaluative aspects. Figure 6b provides a visual representation of these effects, indicating that knowledge-based users were more satisfied if they had higher knowledge levels and vice versa for users of content-based recommenders. We examined whether the path towards choice satisfaction was fully mediated by perceived effort and choice difficulty, but we found no support: $\beta = .090, p = .063$.

4.2.1. Conclusion

We examined the evaluative effects of the interplay between the preference elicitation (PE) method (i.e., content-based or knowledge-based) and user knowledge, operationalized as subjective food knowledge. In doing so, we also explored the interaction effects between labels and user knowledge. Our results revealed two ways in which choice satisfaction was affected. First, users facing MTL labels tended to be more satisfied with their choices if they had a higher knowledge level, while satisfaction levels were slightly lower on average for MTL. On the one hand, this showed the importance of considering domain knowledge in the evaluation of a food recommender system, similar to [10]. On the other hand, it showed that although users of MTL

labels seemed to make slightly healthier choices (but not significantly so), it might have come at the cost of choice satisfaction.

The second main finding concerned the interplay between subjective food knowledge and the PE method. We found that both the perception and the experience of using the system depended on this interaction effect. Users with higher levels of domain knowledge evaluated a recommender more positively if it allowed them to disclose personal characteristics and feature-based preferences, as was done for the knowledge-based PE. In contrast, users with comparatively low levels of domain knowledge evaluated the recommender more positively when facing a content-based method. Although the fully mediated path method was significant, this ‘crossed’ interaction effect was observed for all of the evaluative aspects.

Overall, this study showed that effective personalization in food recommender systems goes beyond algorithmic optimization. In fact, the design of the recommender, in terms of nudges and PE methods, seemed to require careful consideration of the user’s knowledge level. This support the calls for adaptive preference elicitation methods [10].

5. Discussion

For many years, food recommender systems have fallen in line with traditional recommender approaches, focusing on algorithmic optimization. This paper has built upon research in which the interaction methods and interface aspects of a recommender are adapted to support specific user choices [13]. We have singled out the role of a user’s level of domain knowledge, for it may affect not only how recommendations are evaluated, but also what types of interface aspects and interaction methods are appropriate. In doing so, we have focused on the one hand on cognitive food nudges [15], in the form of nutrition labels. On the other hand, we have examined the role of the recommender’s preference elicitation method [10, 13], regarding it as a possible barrier for some users to due to either its complexity (knowledge-based) or simplicity (content-based).

We have set out with two research goals, examining two dimensions of a food recommender system. Firstly, we sought to investigate whether food nutritional labels support recipe recommender users in making healthy food choices. Secondly, we have examined the interplay between subjective food knowledge and preference elicitation methods on the user’s perception and experience in a food recommender system [32], being the first to do so in this domain.

The first main contribution of this paper (RQ1) indicates that annotating personalized recipes with MTL labels or Full labels does not significantly affect the healthiness of recipes chosen by users. This finding falls in line with previous research in food recommender research [27, 22], which found that boosting nutritional food labels is the best way to help users make healthier food choices in a personalized interface, rather than using nudges only. Hence, although digital nudging in recommender systems has gained attention [28], its effectiveness might be limited due to the personalized decision-making context.

Interestingly, our study also reveals that a knowledge-based PE method (and subsequent recommender) leads to healthier outcomes than a more simple content-based recommender. It is possible that simply allowing users to reflect on their own preferences and needs leads to healthier outcomes than relying on recipe ingredients and images only. In the context of the psychological dual-process thinking [29], knowledge-based recommenders might encourage

system-2 thinking, involving extensive and conscious deliberation, while content-based recommenders might elicit intuitive choices (based on system-1). This is consistent with the type of nudges related to effect (e.g., adapting images [18]) and cognition (e.g., labels) [15]. This also highlights the importance of considering user preferences and the interaction method, along with nutritional education, when examining food recommenders.

The second main contribution concerns the interplay between the user's nutritional knowledge and the preference elicitation method. This interaction is shown to influence various evaluative aspects of food recommender systems (RQ2). We find that users with higher levels of food knowledge experience additional benefits when using a knowledge-based recommender and vice versa for a content-based recommender. This has been operationalized into a path model that includes perceived effort, choice difficulty, and choice satisfaction.

Our findings are largely consistent with the work done in the energy recommender system domain [13]. That work differentiates between 'case-based PE' and 'attributed-based PE', which we regard to be similar to our content-based and knowledge-based approaches, respectively. Where both case-based PE and content-based PE (dis)like individual recipes, attribute-based PE differs slightly from a knowledge-based recommender. In [10, 13], users had to make tradeoffs between different energy-saving measure features but did not disclose any personal characteristics. In a knowledge-based recommender, the interactions that involve disclosing needs or personal preferences might have been easier to do, compared to some attribute-based tradeoffs. Another difference is that we have not been able to relate the FSA score in our path model to other aspects, while Knijnenburg et al. [13] also observe a relation between choice satisfaction and the interaction metric 'kWhs saved'.

Another striking finding from our structural equation model is that the evaluation of different labels depends on user knowledge. Where we expected this to be strongest for the full label due to its complexity, we have observed a positive interaction between domain knowledge and the use of an MTL label on choice satisfaction. It seems that the comparatively simple front-of-package label [16] still requires a significant level of understanding to be used satisfactorily.

5.1. Limitations and Future Work

A few limitations might confound parts of this paper. Unfortunately, we have had to exclude around 15% of our participants due to one or more issues regarding non-attention. The fact that this group is rather large could suggest that more users have not engaged with the recommended content with much deliberation. Intuition-based decisions could have undermined the effectiveness of our cognitive labeling nudge [15]. Nonetheless, since we have observed some differences regarding labels in terms of choice and evaluation, we argue that a sufficient number of participants is still part of this study's analysis sample. This also applies to use of Structural Equation Modelling, for which we had a sufficient number of degrees of freedom [41].

The recommender in this study has focused on dinner recipes. Although this is quite a common approach for recipe recommendation [27, 24, 5, 1], it is challenging to assess the implications or a slightly (un)healthier dinner meal if nothing is known about the daily dietary intake of a user. For example, it could be that a person eats relatively healthy dinner meals but has numerous eating moments a day, thereby exceeding the caloric intake limit. We advocate for extending this recommender approach towards meal plans for the day, possibly mixing

recipe recommendations with food product recommendations.

Another limitation, which is shared with many food recommender studies [2], is that we have not checked whether people actually cooked the food. In that sense, the choices made in this study can only be regarded as behavioral intention. Although some commitment mechanisms may take place that may support actual engagement with chosen recipes, we would encourage performing a follow-up study that considers longitudinal aspects as well. An increasing number of health-based personalized advice applications are developed [8], among others using Digital Twins as a means of user profiling to suggest meal plans or exercise behavior. Overall, it is important to examine whether adaptations in a recommender interface can spill over into longer last effects. For instance, a knowledge-based recommender can only be regarded as being effective in supporting healthier choice if these last over the course of a few weeks.

Acknowledgments

This work was supported by industry partners and the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the centers for Research-based Innovation scheme, project number 309339.

References

- [1] C. Trattner, D. Elswailer, Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems, in: Proceedings of the 26th international conference on world wide web, ACM, New York, NY, USA, 2017, pp. 489–498.
- [2] C. Trattner, D. Elswailer, Food recommendations, in: Collaborative recommendations: Algorithms, practical challenges and applications, World Scientific, 2019, pp. 653–685.
- [3] S. Mills, M. White, H. Brown, W. Wrieden, D. Kwasnicka, J. Halligan, S. Robalino, J. Adams, Health and social determinants and outcomes of home cooking: A systematic review of observational studies, *Appetite* 111 (2017) 116–134.
- [4] L. McGowan, M. Caraher, M. Raats, F. Lavelle, L. Hollywood, D. McDowell, M. Spence, A. McCloat, E. Mooney, M. Dean, Domestic cooking and food skills: A review, *Critical reviews in food science and nutrition* 57 (2017) 2412–2431.
- [5] A. D. Starke, E. Asotic, C. Trattner, E. J. Van Loo, Examining the user evaluation of multi-list recommender interfaces in the context of healthy recipe choices, *ACM Transactions on Recommender Systems* (2023).
- [6] D. Elswailer, H. Hauptmann, C. Trattner, Food recommender systems, in: F. Ricci, L. Rokach, B. Shapira (Eds.), *Recommender Systems Handbook*, Springer US, New York, NY, 2022, pp. 871–925. URL: https://doi.org/10.1007/978-1-0716-2197-4_23.
- [7] J. Freyne, S. Berkovsky, Intelligent food planning: personalized recipe recommendation, in: Proceedings of the 15th international conference on Intelligent user interfaces, ACM, New York, NY, USA, 2010, pp. 321–324.
- [8] M. J. Reinders, A. D. Starke, A. R. Fischer, M. C. Verain, E. L. Doets, E. J. Van Loo, Determinants of consumer acceptance and use of personalized dietary advice: A systematic review, *Trends in Food Science & Technology* (2022).

- [9] T. N. T. Tran, M. Atas, A. Felfernig, M. Stettinger, An overview of recommender systems in the healthy food domain, *Journal of Intelligent Information Systems* 50 (2018) 501–526.
- [10] B. P. Knijnenburg, M. C. Willemsen, Understanding the effect of adaptive preference elicitation methods on user satisfaction of a recommender system, in: *Proceedings of the third ACM conference on Recommender systems*, 2009, pp. 381–384.
- [11] D. Elswailer, C. Trattner, M. Harvey, Exploiting food choice biases for healthier recipe recommendation, in: *Proceedings of the 40th international acm sigir conference on research and development in information retrieval*, 2017, pp. 575–584.
- [12] A. Starke, M. Willemsen, C. Snijders, Promoting energy-efficient behavior by depicting social norms in a recommender interface, *ACM Transactions on Interactive Intelligent Systems (TiiS)* 11 (2021) 1–32.
- [13] B. Knijnenburg, M. Willemsen, R. Broeders, Smart sustainability through system satisfaction: tailored preference elicitation for energy-saving recommenders, in: *20th Americas Conference on Information Systems (AMCIS 2014)*, August 7-9, 2014, Savannah, Georgia, United States, AIS/ICIS, 2014, pp. 1–15.
- [14] C. Musto, C. Trattner, A. Starke, G. Semeraro, Towards a knowledge-aware food recommender system exploiting holistic user models, in: *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization*, ACM, New York, NY, USA, 2020, pp. 333–337.
- [15] R. Cadario, P. Chandon, Which healthy eating nudges work best? a meta-analysis of field experiments, *Marketing Science* 39 (2020) 465–486.
- [16] M. Egnell, I. Boutron, S. Péneau, P. Ducrot, M. Touvier, P. Galan, C. Buscail, R. Porcher, P. Ravaud, S. Hercberg, et al., Randomised controlled trial in an experimental online supermarket testing the effects of front-of-pack nutrition labelling on food purchasing intentions in a low-income population, *Bmj Open* 11 (2021) e041196.
- [17] K. L. Hawley, C. A. Roberto, M. A. Bragg, P. J. Liu, M. B. Schwartz, K. D. Brownell, The science on front-of-package food labels, *Public health nutrition* 16 (2013) 430–439.
- [18] A. D. Starke, M. C. Willemsen, C. Trattner, Nudging healthy choices in food search through visual attractiveness, *Frontiers in Artificial Intelligence* 4 (2021) 20.
- [19] S. Barko-Sherif, D. Elswailer, M. Harvey, Conversational agents for recipe recommendation, in: *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, 2020, pp. 73–82.
- [20] G. Castiglia, A. El Majjodi, F. Calò, Y. Deldjoo, F. Narducci, A. Starke, C. Trattner, Nudging towards health in a conversational food recommender system using multi-modal interactions and nutrition labels (2022).
- [21] C. Ischen, T. B. Araujo, H. A. Voorveld, G. Van Noort, E. G. Smit, Is voice really persuasive? the influence of modality in virtual assistant interactions and two alternative explanations, *Internet Research* (2022).
- [22] A. Starke, A. El Majjodi, C. Trattner, Boosting health? examining the role of nutrition labels and preference elicitation methods in food recommendation, in: *Interfaces and Human Decision Making for Recommender Systems 2022*, 2022, pp. 67–84.
- [23] M. Elahi, M. Ge, F. Ricci, D. Massimo, S. Berkovsky, Interactive food recommendation for groups, in: *Poster Proceedings of the 8th ACM Conference on Recommender Systems, RecSys 2014*, Foster City, Silicon Valley, CA, USA, October 6-10, 2014, 2014.

- [24] C. Musto, A. D. Starke, C. Trattner, A. Rapp, G. Semeraro, Exploring the effects of natural language justifications in food recommender systems, in: *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 2021, pp. 147–157.
- [25] N. J. Temple, Front-of-package food labels: A narrative review, *Appetite* 144 (2020) 104485.
- [26] M. W. Kreuter, L. K. Brennan, D. P. Scharff, S. N. Lukwago, Do nutrition label readers eat healthier diets? behavioral correlates of adults' use of food labels, *American journal of preventive medicine* 13 (1997) 277–283.
- [27] A. El Majjodi, A. D. Starke, C. Trattner, Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender system, in: *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, 2022, pp. 48–56.
- [28] M. Jesse, D. Jannach, Digital nudging with recommender systems: Survey and future directions, *Computers in Human Behavior Reports* 3 (2021) 100052.
- [29] D. Kahneman, *Thinking, fast and slow*, Macmillan, 2011.
- [30] A. Berge, V. Velle Sjøen, A. D. Starke, C. Trattner, Changing salty food preferences with visual and textual explanations in a search interface, in: *HEALTHI'21: Joint Proceedings of ACM IUI 2021 Workshops*, CEUR, 2021.
- [31] Department of Health and Social Care UK, *Front of Pack nutrition labelling guidance*, 2016. URL: <https://www.gov.uk/government/publications/front-of-pack-nutrition-labelling-guidance>.
- [32] B. P. Knijnenburg, M. C. Willemsen, Evaluating recommender systems with user experiments, in: *Recommender systems handbook*, Springer, 2015, pp. 309–352.
- [33] C. Trattner, D. Elsweiler, An evaluation of recommendation algorithms for online recipe portals, in: *Proceedings of the 4th International Workshop on Health Recommender Systems*, CEUR, Aachen, DE, 2019.
- [34] J. Oh, K. Yun, J.-H. Chae, T.-S. Kim, Association between macronutrients intake and depression in the united states and south korea, *Frontiers in psychiatry* 11 (2020) 207.
- [35] C. N. Sutanto, M. X. Wang, D. Tan, J. E. Kim, Association of sleep quality and macronutrient distribution: a systematic review and meta-regression, *Nutrients* 12 (2020) 126.
- [36] S. Moats, C. W. Sutor, A. L. Yaktine, S. P. Murphy, et al., *Child and adult care food program: Aligning dietary guidance for all* (2011).
- [37] L. R. Flynn, R. E. Goldsmith, A short, reliable measure of subjective knowledge, *Journal of business research* 46 (1999) 57–66.
- [38] Z. Pieniak, J. Aertsens, W. Verbeke, Subjective and objective knowledge as determinants of organic vegetables consumption, *Food quality and preference* 21 (2010) 581–588.
- [39] A. Starke, M. Willemsen, C. Snijders, Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system, in: *Proceedings of the eleventh ACM conference on recommender systems*, 2017, pp. 65–73.
- [40] M. C. Willemsen, M. P. Graus, B. P. Knijnenburg, Understanding the role of latent feature diversification on choice difficulty and satisfaction, *User Modeling and User-Adapted Interaction* 26 (2016) 347–389.
- [41] J. C. Westland, Lower bounds on sample size in structural equation modeling, *Electronic commerce research and applications* 9 (2010) 476–487.