



Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

A survey-based priority elicitation protocol for community-based resource allocation decisions

 José Geraldo Vidal Vieira^a, Gilberto Montibeller^{b,c,*}
^a Department of Production Engineering, Federal University of São Carlos (UFSCar), Brazil. Address: UFSCar - Campus Sorocaba, Rodovia João Leme dos Santos (SP-264), Km 110 - Bairro do Itinga, Sorocaba - SP, CEP 18052-780, Brazil

^b University of Bristol Business School, University of Bristol, UK (corresponding author). Address: 11-13 Tyndall's Park Rd, Bristol, BS8 1PY, UK

^c KFUPM Business School, Academic Belt Road, KFUPM, Dhahran 31261, Saudi Arabia

ARTICLE INFO

Keywords:

Behavioural operational research

Community OR

Priority elicitation

Swing weighting

Pairwise comparison

Multi-attribute value theory

ABSTRACT

Decision Analysis is increasingly being used to support resource allocation in communities, given its ability to represent the priorities of community members and support transparent resource allocation. Current attempts to elicit priorities in these interventions rely increasingly upon conducting surveys (face-to-face or online). Nevertheless, such preference elicitation initiatives may suffer if respondents do not clearly understand the questions being asked. Many protocols for priority elicitation currently used in Decision Analysis were originally designed to be employed by a decision analyst, who provides extensive support to a small number of decision makers in eliciting their judgments. However, such standard elicitation protocols may not be suitable for surveys, as the elicitation questions require a high level of understanding and a high cognitive effort from the respondents. Hence, in this paper, we suggest a new protocol for eliciting individual priorities for resource allocation decisions via either assisted or unassisted large-scale surveys, which elicits strict preference relations. We base this protocol on the Marketing research literature, which has dealt extensively with similar surveys. We adopt a Multi-Attribute Value Theory framework and design the protocol to avoid the range-insensitivity bias in multi-attribute choices. We assess the suitability of the widely employed swing weighting method for survey-based elicitation of priorities in comparison to the proposed protocol and find that swing weighting may not be suitable for resource allocation problems. We also suggest how the proposed protocol may improve the coherence of judgments elicited from the swing weighting method for survey-based priority elicitation.

1. Introduction

Policymakers at different levels, including municipalities, state governments, and NGOs, face an ever-challenging decision: determining the allocation of scarce resources across diverse community needs (such as health, education, security, sports, transportation, and more). Resource allocation decisions are always complex, involving multiple objectives, several strategic alternatives, and many community stakeholders who may pursue different interests (Montibeller et al., 2009; Haag et al., 2019; Lienert et al., 2016).

In Operational Research (OR), the most common methods to support community-based resource allocation decisions are Portfolio Decision Analysis (Hummel et al., 2017; Liesiö et al., 2007, 2020; Salo et al., 2011) and Multi-Criteria Decision Analysis (Bana e Costa, 2001; Lienert et al., 2016; Montibeller et al., 2009). In both methods, preferences must

be elicited from community members to represent marginal value over attributes and define the relative value of these attributes in the multi-criteria model. We focus this paper on the latter parameter, *priority elicitation*, given how critical the adequate prioritisation of objectives is for guiding the allocation of resources from the perspective of the community.

Decision analysts have often adopted a participative approach (Phillips & Bana e Costa, 2007) and facilitated decision modelling (Franco & Montibeller, 2010) when supporting resource allocation decisions. These Decision Analysis (DA) interventions typically involve a small group (around 5 to 15 members) of decision makers (Eden, 1992; Phillips, 2007). These decision makers are extensively supported by a decision analyst who elicits their preferences for the allocation of resources. However, if this approach is employed to support community resource allocation decisions, this small group may not fully represent

* Corresponding author.

E-mail addresses: jose-vidal@ufscar.br (J.G.V. Vieira), g.montibeller@bristol.ac.uk (G. Montibeller).

<https://doi.org/10.1016/j.ejor.2025.07.049>

Received 11 February 2024; Accepted 21 July 2025

Available online 22 July 2025

0377-2217/© 2025 Published by Elsevier B.V.

the community's preferences or may promote their own agendas instead of those of the community they are supposed to represent.

A more recent trend in community OR is to employ online surveys to elicit the preferences of community members (Aubert et al., 2020, 2022, 2023, 2024; Aubert & Lienert, 2019; Haag et al. 2022; Kuller et al., 2022; Lienert et al., 2016; Scholten et al., 2015; Zheng & Lienert, 2018). These online surveys typically operationalise standard DA preference elicitation methods for attribute weighting, such as the widely employed swing weighting method. The survey elicits preferences from hundreds of community members (Aubert et al., 2020, 2022, 2023, 2024; Aubert & Lienert, 2019; Haag et al. 2022; Kuller et al., 2022; Lienert et al., 2016), thus mitigating the drawbacks of small-group interventions. However, community members may find these elicitation methods challenging (Danielson & Ekenberg, 2019; Rezaei, 2021), particularly in deprived communities, which may impact the coherence of their responses.

This paper proposes a new protocol for eliciting individual priorities for resource allocation decisions through either assisted or unassisted large-scale surveys, given the potential benefits of this elicitation mode for community OR and the challenges associated with using standard elicitation protocols from Decision Analysis. The *assisted mode* involves a survey interviewer who guides the respondent through the elicitation protocol that requires binary choices of competing resources employing cards. The *unassisted mode* is designed to allow respondents to provide the same types of preference information independently, using the written instructions provided, with the elicitation protocol requesting binary choices for competing resources.

We base the suggested protocol on the Marketing research literature, which has extensively dealt with similar surveys, but conceptualised it within a Multi-Attribute Value Theory (MAVT) framework (Keeney & Raiffa, 1993), and adapted the protocol to avoid the range-insensitivity bias in multi-attribute choices. We also analyse a widely employed elicitation protocol for preference elicitation in resource allocation decisions, the swing weighting method, and experimentally assess its suitability for eliciting preferences in surveys against the proposed protocol.

Our contributions are threefold. First, we propose a streamlined and easily implementable protocol designed to elicit individual priorities for resource allocation decisions through surveys. This protocol incorporates best practices from Decision Analysis, such as anchoring judgments on the ranges of the attributes for eliciting preferences. Second, we examine the suitability of the swing weighting method for survey-based elicitation, shedding light on its limitations for this elicitation mode. We identified some weaknesses in the application of this popular elicitation method to survey-based preference elicitation in resource allocation decisions. Third, to address this issue with the swing weighting method, the suggested protocol could also be used to screen out respondents with intransitive preferences, thereby potentially improving the coherence of elicited judgments from swing weighting.

The remainder of the paper is organised as follows. Section 2 reviews the survey-based preference elicitation used in Decision Analysis and Marketing. Section 3 introduces the survey-based preference elicitation protocol, Section 4 presents the research questions and propositions for testing the protocol, and Section 5 describes the research design we adopted to test it. Section 6 presents the experimental results, and Section 7 discusses the findings of our experiments. Section 8 draws our conclusions and suggestions for further research avenues.

2. Survey-Based preference elicitation in decision analysis and marketing

Two concepts, "preference elicitation" and "choice behaviour", are used interchangeably in the OR literature (Falk et al., 2022; Kimbrough & Weber, 1994) and in the Marketing literature (Aribarg et al., 2017). However, these concepts do have different meanings. The former concept involves differentiation among attributes that add utility or

value to the decision alternatives (Huber et al., 1993; Keeney, 1988). The latter concept refers to the "choice between alternatives" (Luce, 1959), which can typically be estimated by the relative frequency of choices from a series of questions designed to elicit preferences regarding alternatives and levels of attributes within repetitive tasks (Tversky, 1972). Given our focus on survey-based elicitation protocols for supporting resource allocation decisions, we cover only the literature on preference elicitation that is relevant to our context in the fields of Decision Analysis and Marketing. Hence, the related literature in Decision Theory on preference modelling (see Keeney and Raiffa (1993)) and in Artificial Intelligence on preference learning (see Fürnkranz and Hüllermeier (2010)) are excluded from this brief review.

2.1. Survey-based priority elicitation in the decision analysis literature

The trade-off method is the most robust theoretical foundation for attribute-weight elicitation (Keeney & Raiffa, 1976). However, this method requires attributes to have continuous measurement scales for identifying equal-value points (Keeney, 1988). In practice, the questions are somewhat challenging to answer (Borcherding et al., 1991), and eliciting weights becomes operationally complex and cognitively demanding due to the substantial number of value judgments required (Riabacke et al., 2012).

Another method widely employed in standard DA practice is the swing weighting (SW) method, which is relatively simple and sound. In this elicitation protocol, the decision maker (DM) must compare and value a change (or swing) from the least-preferred to the most-preferred level on one attribute to the swing from the least-preferred to the most-preferred level on another attribute (Moshkovich et al., 2002; von Winterfeldt & Edwards, 1986; Weber & Borcherding, 1993).

While there are several other attribute weight elicitation methods in standard DA (Belton & Stewart, 2002; Riabacke et al., 2012), most survey-based decision analysis interventions have employed the SMART method (von Winterfeldt & Edwards, 1986), the Deck of Cards method (Figueira & Roy, 2002), or (online) SW protocols (Keeney & Raiffa, 1976). The preference protocol of SMART for attribute importance does not consider attribute ranges, so it is not recommended. The Deck of Cards method requires the ordering of attributes and can be employed for a multi-attribute value analysis with the definition of attribute ranges (Corrente et al., 2021; Figueira & Roy, 2002; Pictet & Bollinger, 2008). The protocol is relatively straightforward in terms of attribute ordering. However, the method requires the assessment of quantitative ratio judgments to determine attribute weights, a cognitively demanding task (Larichev, 1992) that may be challenging to perform for large-scale surveys.

As mentioned above, the SW method has been amply and successfully employed in standard DA interventions. However, a decision analyst must guide and support the decision maker during the preference elicitation phases in those interventions. This is not the case of survey-based elicitation processes, in which there is either a mere interviewer with somewhat limited knowledge of DA or a self-report form (online or on paper).

There is some evidence of challenges in adopting the SW method in (online) survey-based preference elicitation (in contrast to other elicitation protocols in standard DA analysis that have not been employed online and may prove even more challenging in this type of intervention). First, it has been consistently observed that respondents often exhibit low self-reported confidence in their elicited preferences (Aubert et al., 2024). Second, several reports identify that it was challenging for survey participants to understand the elicitation process (Aubert et al., 2020, 2024; Haag et al. 2022; Lienert et al., 2016), despite detailed instructions as well as assistance during the procedure (Danielson & Ekenberg, 2019). Third, there are indications of process compliance failures, as respondents do not correctly follow the instructions of the elicitation survey when using this method in preference elicitation surveys (Aubert et al., 2020; Danielson & Ekenberg, 2019). Therefore, it is

an open research question whether this elicitation protocol is suitable for survey-based preference elicitation in resource allocation decisions.

2.2. Survey-based preference elicitation in the marketing literature

The fields of Marketing and Consumer Behaviour have developed many protocols for eliciting individual judgments of multi-attribute preferences for products and services (Batsell, 1980; Louviere, 1988; Carson et al., 1994; Rao, 2014). Despite these sophisticated methods for large-scale preference surveys, their elicitation protocols often focus on choice or lack rigour from a DA perspective (e.g., most do not consider ranges of attributes when eliciting attribute weights). There are several methods in Marketing based on paired comparisons (Scholz et al., 2010) and self-explicated approaches (Schlereth et al., 2014), which are relevant in our context, as their elicitation protocols are similar to those employed for attribute weight elicitation in DA. In these methods, once the respondent selects the most valuable attribute (the anchor), they evaluate the remaining attributes using a rating or ranking scale. This anchor must be explicit (Payne et al., 1992), making the elicitation process easier for respondents (Srinivasan, 1988).

A particularly relevant method for our decision context is the paired comparison method (Scholz et al., 2010). In this method, respondents evaluate a pair of profiles, either showing alternatives described by their performance on every attribute (full profile) or else described by only two attributes (partial profile) (Rao, 2014). The preference between the two profiles is elicited either by a Likert scale or by determining which attribute is most preferred. According to Rao (2014), the advantage of the two (partial) profiles is that the respondent is asked to focus on two product/service concepts; therefore, the evaluations may be more meaningful. The pairwise comparison (with full or partial profiles) is most effective when the number of alternatives involved is small (less than seven) (Larichev, 1992; Larichev & Moshkovich, 1995), the range of performances for the relevant alternatives involved is not wide, and the decision maker has a clear understanding of the relative value of the alternatives (Ngwenyama & Bryson, 1999).

Summarising the existing literature, in Decision Analysis, most priority elicitation methods were developed for a decision analyst to elicit preferences from a small number of decision makers. The emerging Decision Analysis literature based on survey-based preference elicitation is still limited and has indicated challenges in applying standard elicitation protocols (such as the SW method) in this context. On the other hand, in Marketing, approaches for preference elicitation were developed for large-scale surveys, but with limited consideration of the axiomatic aspects of preference elicitation required in Decision Analysis.

3. The proposed survey-based preference elicitation protocol

In this section, we suggest a new Marketing-based protocol for survey-based preference elicitation in resource allocation decisions. This novel approach may enable policymakers in community-based settings to understand the preferences of individual community members over resource improvements.

3.1. The conceptual background of the proposed elicitation protocol

There is some evidence that decision makers often resist sophisticated elicitation methods (Kottemann & Davis, 1991) and prefer more straightforward and easier-to-understand methods, such as pairwise comparisons (Zheng & Lienert, 2018). Because some respondents are easily distracted while rating or scoring alternatives (Bryson et al., 1995), pairwise comparisons may facilitate their choices. In addition, judgments of strict preference are cognitively easier to make and more stable than the quantitative judgments required by the trade-off and SW methods (Larichev, 1992; Montibeller, 2018; Olson & Dorai, 1992).

On the other hand, a key characteristic of any sound elicitation protocol for the elicitation of attribute weights is the need to anchor the

judgments on the ranges of the attributes (Belton & Stewart, 2002; Montibeller, 2018) to avoid the range-insensitivity bias (von Nitzsch & Weber, 1993). Our proposed protocol implements this requirement. In addition, the suggested protocol is based on the pairwise comparison of partial profiles employed in Marketing research (Green & Wind, 1975), which simplifies the judgment task if compared to full profiles. The judgment task of the protocol requires only judgments of strict preference among pairs of partial profiles.

A significant advantage of employing an adapted protocol from Marketing is that it was originally designed to be used in preference-elicitation surveys, in contrast to the protocols developed in DA, which were intended to be administered by a decision analyst. In addition, we designed the protocol to minimise the cognitive burden, as it does not require any quantitative judgment, only judgments of strict preference. The protocol is formalised next.

3.2. Formalising the survey-based elicitation protocol

The elicitation protocol we suggest is conceptualised within an MAVT framework (Keeney & Raiffa, 1993). Let $\mathbf{X}$ be a set of N (with $N > 1$) attributes, with X_i representing distinctive resource attributes:

$$\mathbf{X} = \{X_1, X_2, \dots, X_N\}$$

The levels of each attribute describe ordered improvements of the respective resource, with the lower bound denoting the status quo situation and the upper bound denoting the best possible situation. Hence, for each X_i attribute we define two bounds:

$$X_i^0 = \text{the lower bound of the } i\text{-th attribute}$$

$$X_i^* = \text{the upper bound of the } i\text{-th attribute}$$

Respondents are asked to compare two profiles. The first profile (X_i^*, X_j^0) , with $i \neq j$, represents a dummy alternative in which the i -th resource is at the best level (hypothetical situation) and the j -th resource is at the status quo level. The second profile (X_j^*, X_i^0) , again with $i \neq j$, represents a dummy alternative in which the j -th resource is at the best level (hypothetical situation) and the i -th resource is at the status quo level. The respondent is then asked to establish their strict preferences over these two profiles, selecting one of the two dummy alternatives:

$$(X_i^*, X_j^0) \succ (X_j^*, X_i^0) \text{ OR } (X_j^*, X_i^0) \succ (X_i^*, X_j^0)$$

with $i \neq j$ (and $i, j = 1, 2, \dots, N$)

The attributes must fulfil the preferential independence condition (see Keeney (1992)). Specifically, the strict preference between a pair of profiles does not depend on the levels at which the remaining t -th attributes are fixed. The profiles assume that such levels are set at the status quo level, i.e. X_t^0 of each resource (with $t \neq i, j$ and $t = 1, 2, \dots, N$). The method requires $N(N-1)/2$ pairwise comparisons.

For example, two attributes could be represented by the following types of resources for the community: Health system (X_{HS}) and Education system (X_{ES}). By choosing one of the profiles (X_{ES}^*, X_{HS}^0) or (X_{HS}^*, X_{ES}^0) , the respondent indicates the preferred resource improvement.

These judgments between every two profiles establish an individual ordinal ranking of the resources by each respondent. A key logical property of such a ranking is its transitivity. The transitivity of the ranking of resource profiles has been defined as follows. Consider three attributes X_i, X_j, X_k , with three profiles (X_i^*, X_j^0) , (X_i^0, X_j^*) and (X_j^*, X_k^0) . If the respondent prefers the first to the second profile, i.e. $(X_i^*, X_j^0) \succ (X_i^0, X_j^*)$, and the second to the third profile, i.e. $(X_i^0, X_j^*) \succ (X_j^*, X_k^0)$,

(X_j^0, X_k^*) , then the ranking is transitive if and only if the respondent prefers the first to the third profile, i.e.: $(X_i^*, X_k^0) \succ (X_i^0, X_k^*)$.

4. Research questions and propositions

In this section, we detail the two research questions that guided the design of our behavioural experiment, which are linked to the proposed elicitation protocol.

RQ 1: *Is the Swing Weighting method suitable for survey-based preference elicitation in resource allocation decisions?*

Despite being widely employed in practice, there is some evidence of potential weaknesses of the SW method, as mentioned previously. First, several reports of real-world surveys indicate that survey participants faced challenges in understanding the elicitation process required by the SW method (Aubert et al., 2020; Haag et al. 2022; Aubert et al., 2019; Danielson & Ekenberg, 2019). These reports also highlighted process compliance failures. For instance, in one intervention, most participants struggled to follow basic instructions, resulting in a decline in process compliance (Aubert et al., 2020). In another study with 39 respondents, only four clearly understood the difference between absolute and relative weights (Danielson & Ekenberg, 2019). Second, there has been no consensus on the convergent validity between the SW method and other Decision Analysis elicitation methods (e.g., the trade-off method) across different contexts. On one hand, van Ittersum et al. (2007) reported convergent validity (i.e. whether different measurements are positively correlated) among SW and several methods. Indeed, Fischer (1995) and Borcherting et al. (1991) found high convergent validity between SW and the trade-off method. On the other hand, Stewart and Ely (1984) and Pöyhönen and Hämäläinen (2001) found the opposite result. Third, the SW method lacks an internal consistency check when evaluating the consistency of a given preference (Liang et al., 2022; Rezaei, 2021; Rezaei et al., 2022) unless this is ensured by a decision analyst conducting the individual preference elicitation (von Winterfeldt & Edwards, 1986). Fourth, multiple reports consistently highlighted range-insensitivity bias (Montibeller & von Winterfeldt, 2015) in the SW method (Liang et al., 2022; Fischer, 1995; Borcherting et al., 1991; Stewart & Ely, 1984). The proposed protocol provides an opportunity to investigate RQ1, by cross-checking the responses from the SW method with strict preference relations for survey-based preference elicitation of the proposed protocol.

RQ2: *To what extent can the proposed protocol for survey-based preference elicitation provide coherent responses in resource allocation decisions?*

We are interested in testing our protocol, which leads to a second research question (RQ2). To address these two research questions, we propose the following six propositions. We tested these propositions by employing two treatments: pen-and-paper (unassisted mode) and ruler-and-cards (assisted mode), which are detailed in the next section. The first four propositions address RQ1, while the last two address RQ2.

- P₁: Process compliance failures occur in the SW method for survey-based preference elicitation (in the assisted and the unassisted modes).
- P₂: The consistency between the SW method for survey-based preference elicitation and the ordinal ranking of preferences elicited from the proposed protocol is low (in the assisted and the unassisted modes).
- P₃: The first swing (which anchors the value scale) and the last swing on resources elicited from the SW method for survey-based preference elicitation show low consistency with the ordinal

ranking of preferences elicited from the proposed protocol (in the assisted and the unassisted modes).

- P₄: The range of weights produced by the SW method for survey-based preference elicitation may be influenced by the treatment (pen-and-paper or ruler-and-cards) employed in its elicitation.
- P₅: For the assisted mode, the survey interviewer influences:
 - P_{5.1}. The transitivity in the resource attribute ranking in the proposed protocol.
 - P_{5.2}. The consistency between the SW method for survey-based preference elicitation and the ordinal ranking of preferences elicited from the proposed protocol
- P₆: For the assisted mode, the application sequence (ruler-and-cards versus cards-and-ruler) influences:
 - P_{6.1}. The transitivity in the resource attribute ranking in the proposed protocol.
 - P_{6.2}. The consistency between the SW method for survey-based preference elicitation and the ordinal ranking of preferences elicited from the proposed protocol.

5. Research design

This section outlines the research design used to test the survey-based preference elicitation protocol and to compare it with the SW method for survey-based preference elicitation in the resource allocation interventions.

5.1. Experimental design

Respondents. Our sample comprises 266 subjects (voluntary students) at a public university in Brazil who participated in the survey between May and July 2023. We selected students from Economics and Production Engineering programmes, due to their basic knowledge about Microeconomics (Preference/Utility Theory). In total, 68 % of the sample were Engineering students and 32 % were Economics students; 63 % identified themselves as male and 27 % as female.

Questionnaire design. The questionnaire consists of two sections. For the first section, we selected the explanatory variables of gender, programme, and number of semesters studied in the programme. The second section includes questions for eliciting preferences in the resource allocation decision: the SW procedure and the proposed protocol. The final version of the questionnaire is presented next (see the supplementary material for details about the pre-testing phase).

Resource Attributes. Table 1 presents each resource attribute and its two levels: the current status quo level and the potential improvement level. The aim was to select the resources that students experience in their daily lives on campus. We opted for a realistic setting as individual preferences from lab-based (consumer) experiments might not translate directly to real decision-making processes (Kuller et al., 2022; Danielson & Ekenberg, 2019). These resources and respective status quo levels were designed based on a workshop group organised by undergraduate students of the third and fourth academic years of Engineering

Table 1

Resource attributes and their respective *status quo* and *improvement* levels.

Resource	Attribute level – status quo	Attribute level – improvement
Sports facilities (SF)	One outdoor court	Indoor athletics centre, outdoor stadium, gym, swimming pool, toilets, etc.
Leisure (Le)	Limited infrastructure for resting and leisure	Thematic areas for leisure, rest; and minimarket.
Laboratory (Lab)	Limited access and a few computers	Expanded access, new computers, print server.
Library (Lib)	Reduced timetable, limited collection and few study rooms	Extended hours, including Saturdays, expanded collection and multiple study rooms.
Restaurant (Res)	University restaurant	University restaurant, facilities with several options for lunch

and Economics. After that, during two online meetings between the students' leaders and the researchers, the subject completed the possible improvement levels that the university could provide with future resource investments.

Procedure. The procedure involves two treatments: (1) pen-and-paper via a survey (unassisted mode) and (2) ruler-and-cards with a survey interviewer (assisted mode by a survey interviewer). The subjects were randomly divided into two treatments, maintaining a balance of Engineering and Economics students and gender. Before starting the survey, the subjects were provided with information about the research objectives and the anticipated time required for participation. A consent statement needed to be agreed upon by each participant, and participants were informed about a token incentive for their participation (a chocolate bar) and a prize draw (a chocolate box) for those who demonstrated consistency between the two tasks. We describe the two treatments next.

Pen-and-paper Treatment: In this unassisted mode, participants were invited to complete the SW procedure adapted for use with a pen-and-paper survey (see supplementary material for a copy of the instructions). A continuous scale ranging from 0 to 10 was used to score the attribute resources, an adapted version of the SW method (Task 1). We used the protocol outlined in Section 3.2 for Task 2, with subjects asked to express strict preferences among pairs of resource attributes. This survey was conducted with participants in the classroom, either during or after classes.

Ruler-and-cards Treatment: In this assisted mode, participants were individually invited to participate either in their classrooms or a suitable room nearby. The SW procedure (Task 1) and the proposed protocol (Task 2) were applied sequentially, with three survey interviewers taking turns. All interviewers received extensive training from one of the authors before applying the survey (during three sessions, simulating the procedures with external subjects, totalling four hours). Each interviewer elicited preferences individually from every subject. (However, the interviewer has not provided the same level of involvement as a decision analyst in a standard facilitated DA intervention.) We present this treatment next (see the supplementary material for a detailed description).

For Task 1, we employed a 0–100 continuous scale (see Fig. 1) and used the graphical method to elicit value scores. As the initial step of SW elicitation, the interviewer showed the resource attributes and their respective *status quo* and *improvement* levels. The survey interviewer used different coloured crayons (represented by a semi-ring in the same figure) to describe the respective resources.

Then, the interviewer described the situation where all resource attributes are at the *status quo* level. This situation was represented at the lower end of the measurement scale; a hypothetical situation for improvements in resource allocations is also presented (see Fig. 1A).

In the first step, the resources at the status quo level are fixed on the scale with a 0 value (see Fig. 1B). The survey interviewer also showed the side of the ruler without the scale to the subjects, presenting it as a continuum scale and only informed them about the upper (100 value) and lower (0 value) limits.

In the second step, the survey interviewer introduced a hypothetical scenario involving resources, where only one resource attribute was at its most preferred level, while all others remained at their current status quo level. The survey interviewer asked the subject to position the most preferred attribute resource swing (first swing) at the top of the ruler with a value of 100 (see Fig. 1C, where the participant's selection of the Restaurant resource swing is indicated in yellow).

In the third step, the subject was asked to select the second most relevant resource attribute swing and position it in terms of its value distance between the top (100 value score) and the bottom of the ruler (0 value score), indicating their choice of the Laboratory resource swing in green in this example. This valuation was repeated for all remaining attribute resources (see Fig. 1D), in order of preference from the most preferred (first swing positioned on the top) to the least preferred (last swing positioned at the bottom), always comparing each swing with the most preferred swing. Subjects completed these evaluations using crayons without seeing the scale on the ruler. Afterwards, the interviewer checked the measurements on the ruler and recorded the respective value scores.

For Task 2, we implemented the proposed elicitation protocol, which involved introducing questions based on a set of trade-off alternatives using profile cards (in a manner similar to the initial criteria ranking in

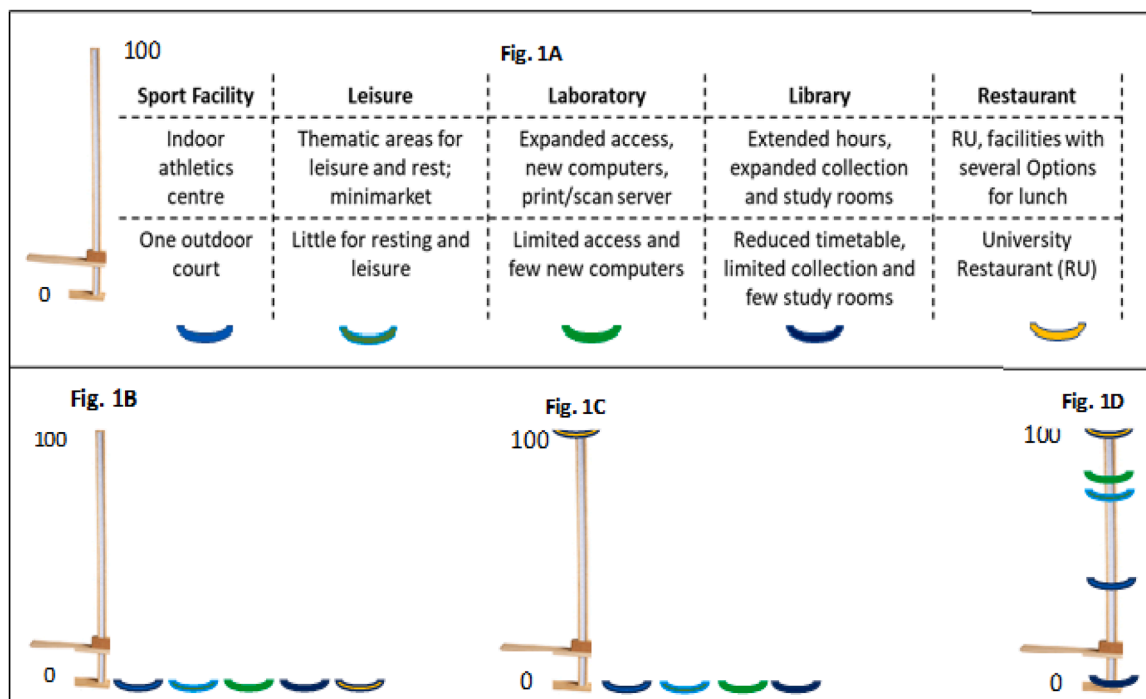


Fig. 1. An example of Swing weight elicitation using the ruler.

the Deck of Cards protocol (Figueira & Roy, 2002; Corrente et al., 2021)). As we have five attributes to be compared in pairs, we performed 10 pairwise comparisons. The survey interviewer presents these cards in pairs, as exemplified in Fig. 2. By choosing between one of the cards, (X_{SF}^*, X_{Lab}^0) or (X_{Lab}^*, X_{SF}^0) , the subject indicates the more valuable resource attribute.

Fig. 3 summarises the research design, with the prioritisation of *Resource Attributes*, the two *Elicitation Tasks* (SW vs the proposed protocol), and the two *Treatments* (pen-and-paper versus ruler-and-cards). Numbers in brackets indicate the number of subjects in each treatment. We also added the related propositions ($P_1, P_2, P_3, P_4, P_{5.1}, P_{5.2}, P_{6.1}$ and $P_{6.2}$) connected with the research design components.

5.2. Variables for the data analysis

The elicited value of attributes resources was measured as quantitative variables on an interval scale: point estimation on a continuum interval scale for Task 1 (SW method) and binary pairwise comparisons for Task 2 (proposed protocol). In addition, the interval scale in the ruler-and-cards treatment employed a 0 to 100 range but, in the data analysis, these scores were divided by 10 to meet the same scale (0–10) as in the pen-and-paper treatment. The checking variables were classified into nominal and ordinal variables, as described in Table 2.

All data were analysed with non-parametric tests as the samples are nominal and ordinal, making parametric tests inappropriate. The data analysis was performed using the IBM Statistical Package for the Social Sciences (SPSS, version 21.0), Python and Microsoft Excel.

5.3. Summary of the research design

Table 3 summarises the research design, showing the propositions, procedure, and data for each treatment.

6. Experimental results

This section outlines the experimental results for the unassisted mode (pen-and-paper treatment) and the assisted mode (ruler-and-cards treatment). Of 143 answers for pen-and-paper treatment, 141 were completed for Task 1, and 140 were completed for Task 2. All subjects ($n = 123$) participating in the second treatment (ruler-and-cards) completed both tasks.

6.1. Process compliance

Proposition 1: Process compliance failures occur in the SW method for survey-based preference elicitation (in the assisted and the unassisted modes).

For the pen-and-paper treatment, the subjects were requested to complete the following two steps using a 0–10 continuum scale. First, after introducing the status quo level and the respective improvement level for each resource attribute, we asked subjects to score 10 for the first swing (where the resource moves from the status quo to the ideal level). Second, the subjects were asked to value the subsequent swings of resource attributes, comparing them with the first swing.

In this treatment, 16.3 % of subjects had compliance failures in

following its instructions (23 of 141 subjects; 95 % CI [10.21 %, 22.41 %]). It means that, in repeated sampling under the same experimental conditions, 95 % of such intervals would contain the true population proportion of the subjects who had compliance failures. Specifically, in the first step, 10 % of our sample chose more than one resource improvement as the best one; 4 % did not rate a score of 10 to any resource improvement; and 2 % completed only the first step or revealed no understanding and did not complete step two. The subjects used only integer scores on the continuum scale, which may indicate an absolute valuation instead of a relative assessment.

No process compliance failures were observed in the ruler-and-cards treatment (assisted mode). Under this latter treatment, all subjects provided non-integer value scores, and none chose more than one resource improvement as the best one.

6.2. Consistency of preferences

Proposition 2: The consistency between the SW method for survey-based preference elicitation and the ordinal ranking of preferences elicited from the proposed protocol is low (in the assisted and unassisted modes).

In our sample, only 33.1 % showed total consistency between Task 1 and Task 2 (88 out of 266 subjects; 95 % CI [27.4 %, 38.7 %]). This CI indicates that, if the sampling process were repeated under identical experimental conditions, 95 % of the resulting confidence intervals would be expected to include the true proportion of subjects with total consistency between Tasks 1 and Task 2. Table 4 shows the consistency levels for all samples and treatments. We conducted a χ^2 test, which indicates that the consistency levels are not equal within each group.

Table 4 also shows the transitivity in Task 2. In the full sample, 79.3 % of the subjects maintained transitivity in Task 2; 78.3 % maintained transitivity in Task 2 for the pen-and-paper treatment and 80.5 % for the ruler-and-cards treatment. If we consider only the subjects with fully transitive preferences in Task 2, disregarding the treatments, only 41.8 % of subjects were totally consistent between tasks. In the pen-and-paper 53.6 % of the subjects were totally or partially consistent between tasks, and 62.6 % were in the ruler-and-cards treatment. In the latter treatment, considering only one change in the resource ranking and including all distances between changes in task 1, the maximum consistency level reached was 68.8 %. This means that 31.32 % of the subjects changed the ranking of at least two resources between tasks.

Fig. 4 presents the results of the ranking order for all resources in Task 1 and Task 2, considering both treatments. The subjects' ranking on the diagonal line showed full consistency of ranking between the two tasks. However, many subjects presented no consistency in their rankings between tasks. In the pen-and-paper treatment, more subjects were inconsistent than in the ruler-and-cards treatment, as illustrated by the wider dispersion of rankings from the diagonal line in the same figure.

The Spearman correlations also show smaller coefficients between resource attributes in the pen-and-paper treatment when compared to the ruler-and-cards treatment for subjects with transitive preferences (Table 5). All Spearman correlations were positive.

Card 1 (X_{SF}^*, X_{Lab}^0)	Card 2 (X_{Lab}^*, X_{SF}^0)
X_{SF}^* = Indoor athletics centre, outdoor stadium, gym, swimming pool, toilets, etc. X_{Lab}^0 = Limited access and a few new computers	X_{Lab}^* = Expanded access, new computers, print/scan server X_{SF}^0 = One outdoor court

Fig. 2. Example of two card descriptions for the Sports Facility (SF) and Laboratory (Lab).

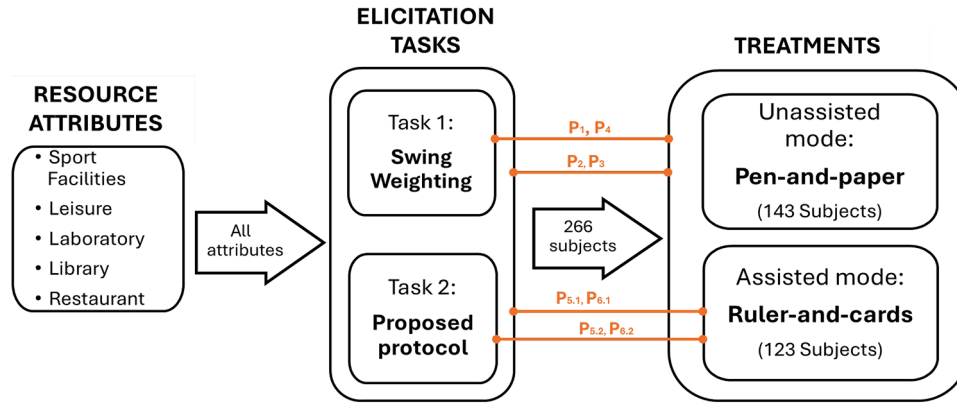


Fig. 3. The research design.

Table 2

Variables for the data analysis.

Variable	Type	Description
Process Compliance ¹	Nominal	0 = No process compliance failure in Task 1; 1 = Process compliance failure in Task 1
Transitivity	Nominal	0 = No transitivity in Task 2; 1 = Transitivity in Task 2
First/Last resource attribute ranking	Nominal	0 = Different (first/last) resource attribute ranking between Task 1 and Task 2. 1 = The same (first/last) resource attribute ranking between Task 1 and Task 2
Consistency level ²	Ordinal	0 = No consistency among resources ranking between Task 1 and Task 2 1 = Partial consistency resources ranking between Task 1 and Task 2 2 = Total consistency resources ranking between Task 1 and Task 2

Notes: ¹ We define a process compliance failure as any mistake in following the instructions required by the Task 1. ² We define “total consistency” when the rank order of the resources in Task 1 is the same in Task 2. In those cases in which total consistency is not achieved, we define “partial consistency” when the rank order of the resources in Task 1 is the same as in Task 2 if the difference in value scores of the resource attribute ranking is within a |1| value point threshold (on a 0 to 10 scale); if the difference in value scores exceeded this threshold, we defined it as “no consistency” between tasks.

Table 3

Summary of propositions, statistical procedures, and data.

Proposition	Procedure	Data
P ₁	Frequency analysis; Confidence Interval	$N_{R-C} = 123$; $N_{P-P} = 143$
P ₂	Confidence Interval; χ^2 test; Spearman correlation	$N_{All} = 266$; $N_{P-P} = 143$; $N_{R-C} = 123$
P ₃	Confidence Interval	$N_{All} = 266$; $N_{P-P} = 143$; $N_{R-C} = 123$
P ₄	Boxplot; Fligner-Killeen test	$N_{P-P} = 143$; $N_{R-C} = 123$
P _{5.1}	Frequency analysis	$N_{R-C} = 123$
P _{5.2}	Kruskal-Wallis test; Cross table analysis	$N_{R-C} = 123$
P _{6.1}	χ^2 test	$*N_{R-C} = 88$; $*N_{C-R} = 35$
P _{6.2}	Mann-Whitney Test	$*N_{R-C} = 88$; $*N_{C-R} = 35$

Notes: N_{All} = All sample; N_{P-P} = Pen-and-paper treatment sample; N_{R-C} = Ruler-and-cards treatment sample.

The data in P₆ (*) represents the sample split into two random groups.

6.3. Consistency among first and last resource attribute ranking

Proposition 3: The first resource attribute ranking (which anchors the value scale) and the last resource attribute ranking on resources elicited from the SW method for survey-based preference elicitation show low consistency

Table 4

Consistency levels for each group (per treatment and in total).

		Pen-and-Paper (%)	Ruler-and-Cards (%)	Total (%)
Transitivity in Task 2		112 (78.3 %)	99 (80.5 %)	211 (79.3 %)
Full sample				
Consistency Levels between Tasks 1 and 2	No consistency	83 (58.0 %)	61 (49.5 %)	144 (54.1 %)
	Partial consistency	14 (9.8 %)	20 (16.3 %)	34 (12.8 %)
	Total consistency	46 (32.2 %)	42 (34.2 %)	88 (33.1 %)
	Sample	143 (100 %)	123 (100 %)	266 (100 %)
	χ^2 (p-value)	50.028 (< 0.01)	20.537 (< 0.01)	68.241 (< 0.01)
Transitive ranking sample				
Consistency Levels between Tasks 1 and 2	No consistency	52 (46.4 %)	37 (37.4 %)	89 (42.2 %)
	Partial consistency	14 (12.5 %)	20 (20.2 %)	34 (16.1 %)
	Total consistency	46 (41.1 %)	42 (42.4 %)	88 (41.8 %)
	Sample	112 (100 %)	99 (100 %)	211 (100 %)
	χ^2 (p-value)	22.357 (< 0.01)	8.061 (< 0.05)	28.161 (< 0.01)

with the ordinal ranking of preferences elicited from the proposed protocol (in the assisted and the unassisted modes).

We analyse the consistency of the resource attribute ranking by comparing the chosen resource attribute ranking in Task 1 versus the resource attribute rankings in Task 2. We used the Confidence Interval test to verify the consistency between the ranking of the first resource attribute and the ranking of the last resource attribute.

Subjects have not always selected the same resource in both tasks, as shown in Table 6. Specifically, the results show that 32.7 % (95 % CI [27.1 %, 38.3 %]) of the subjects selected different first resource attributes and 33.8 % (95 % CI [28.1 %, 39.5 %]) selected different last resource attributes among tasks for both treatments. Considering the treatments separately, for the pen-and-paper, 36.4 % (95 % CI [28.5 %, 44.2 %]) of the subjects selected different first resource attributes and 32.2 % (95 % CI [24.5 %, 39.8 %]) selected different last resource attributes, while for the ruler-and-cards treatment, 28.5 % (95 % CI [20.5 %, 36.4 %]) selected different first resource attributes and 35.8 % (95 % CI [27.3 %, 44.2 %]) selected different last resource attributes.

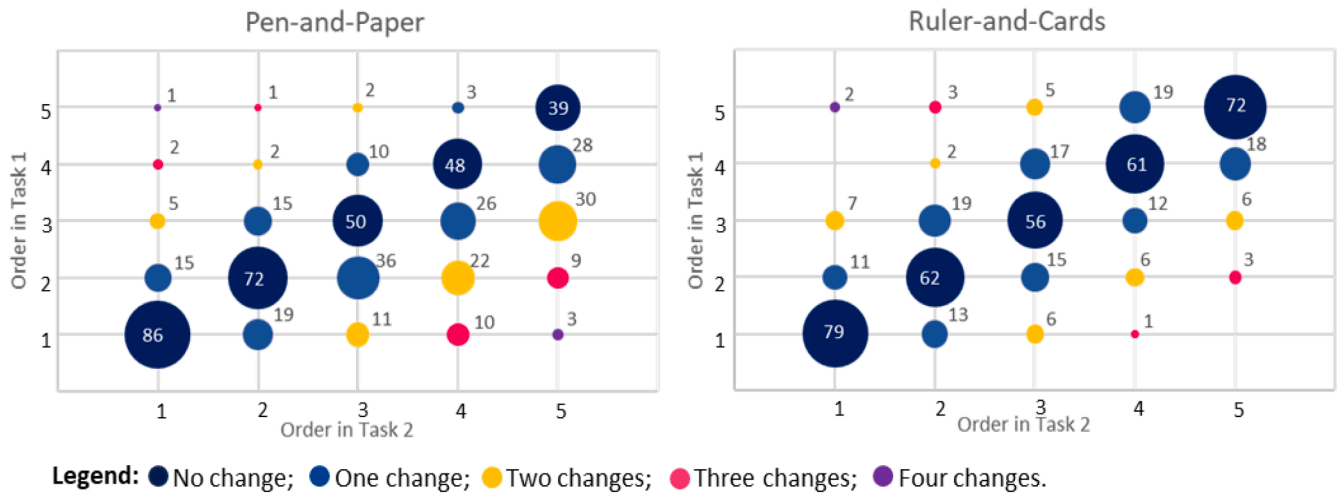


Fig. 4. Order for the resource attributes rankings between tasks in both treatments.

Note: The size of the circles indicates the proportion of rankings in Task 2 and Task 1 (per row).

Table 5

Spearman rank-order correlations between resource attributes in both Tasks.

Pen-and Paper treatment (N = 109)			Ruler-and-Cards treatment (N = 99)		
Order in Task 1	Order in Task 2	ρ	Order in Task 1	Order in Task 2	ρ
Sport	Sport	0.68***	Sport	Sport	0.85***
Leisure	Leisure	0.74***	Leisure	Leisure	0.85***
Laboratory	Laboratory	0.67***	Laboratory	Laboratory	0.75***
Library	Library	0.70***	Library	Library	0.88***
Restaurant	Restaurant	0.74***	Restaurant	Restaurant	0.80***

Note: Significance: *** $p < .001$.

6.4. Resource attribute ranking

Proposition 4: The range of weights produced by the SW method for survey-based preference elicitation may be influenced by the treatment (pen-and-paper or ruler-and-cards) employed in its elicitation.

The distributions of weights in the two treatments, represented with boxplots (Fig. 5), show the range in the pen-and-paper is smaller than the ruler-and-cards treatment (the interquartile range size is larger). On the other hand, there are more outliers in the pen-and-paper treatment than in the ruler-and-cards treatment.

The Fligner–Killeen test was used to assess whether the variances of the resource attributes differ between the two treatments. The results indicated significant difference in variances for $\chi^2_{\text{sport}} = 6.71, p = .00$; $\chi^2_{\text{laboratory}} = 2.83, p = .09$; and $\chi^2_{\text{restaurant}} = 6.20, p = .01$; indicating heterogeneity of variances. For Library and Leisure, the test indicated homogeneity of variances: $\chi^2_{\text{library}} = 0.96, p = .32$; $\chi^2_{\text{leisure}} = 0.16, p = .68$.

6.5. Further analysis of the ruler-and-cards treatment

This subsection analyses the ruler-and-cards treatment for community-based preference elicitation surveys and whether this assisted mode provides more coherent responses than the unassisted mode. To systematically improve this application in communities, we compare the proposed protocol with assisted SW by following two propositions.

Proposition 5. For the assisted mode, the survey interviewer influences:

P_{5.1}. The transitivity in the resource attribute ranking in the proposed protocol.

Table 6

Proportions and Confidence Intervals test for consistency between the first and last resource attribute rankings.

Treatment	First/Last resource attribute ranking	N	Proportions	95 % Confidence Intervals
Pen-and-Paper and Ruler-and-Cards	First resource attribute ranking:	179		
	Equal			
	Different	87	32.7 %	[27.1 %, 38.3 %]
Pen-and-Paper	Last resource attribute ranking:	176		
	Equal			
	Different	90	33.8 %	[28.1 %, 39.5 %]
Pen-and-Paper	Total	266		
	First resource attribute ranking:	91		
	Equal			
	Different	52	36.4 %	[28.5 %, 44.2 %]
Ruler-and-Cards	Last resource attribute ranking:	97		
	Equal			
	Different	46	32.2 %	[24.5 %, 39.8 %]
Ruler-and-Cards	Total	143		
	First resource attribute ranking:	88		
	Equal			
	Different	35	28.5 %	[20.5 %, 36.4 %]
Ruler-and-Cards	Last resource attribute ranking:	79		
	Equal			
	Different	44	35.8 %	[27.3 %, 44.2 %]
Total		123		

P_{5.2}. The consistency between the SW method for survey-based preference elicitation and the ordinal ranking of preferences elicited from the proposed protocol.

The percentage of subjects that maintain transitivity in Task 2 is similar across the interviewers (Interviewer A = 79 %; Interviewer B = 82 %; Interviewer C = 83 %; Crosstable analysis, $p = .001$) (P_{5.1}), while the consistency levels between tasks revealed differences according to the interviewers. We employed the Kruskal–Wallis test to verify if there is a difference in consistency levels among subjects when different

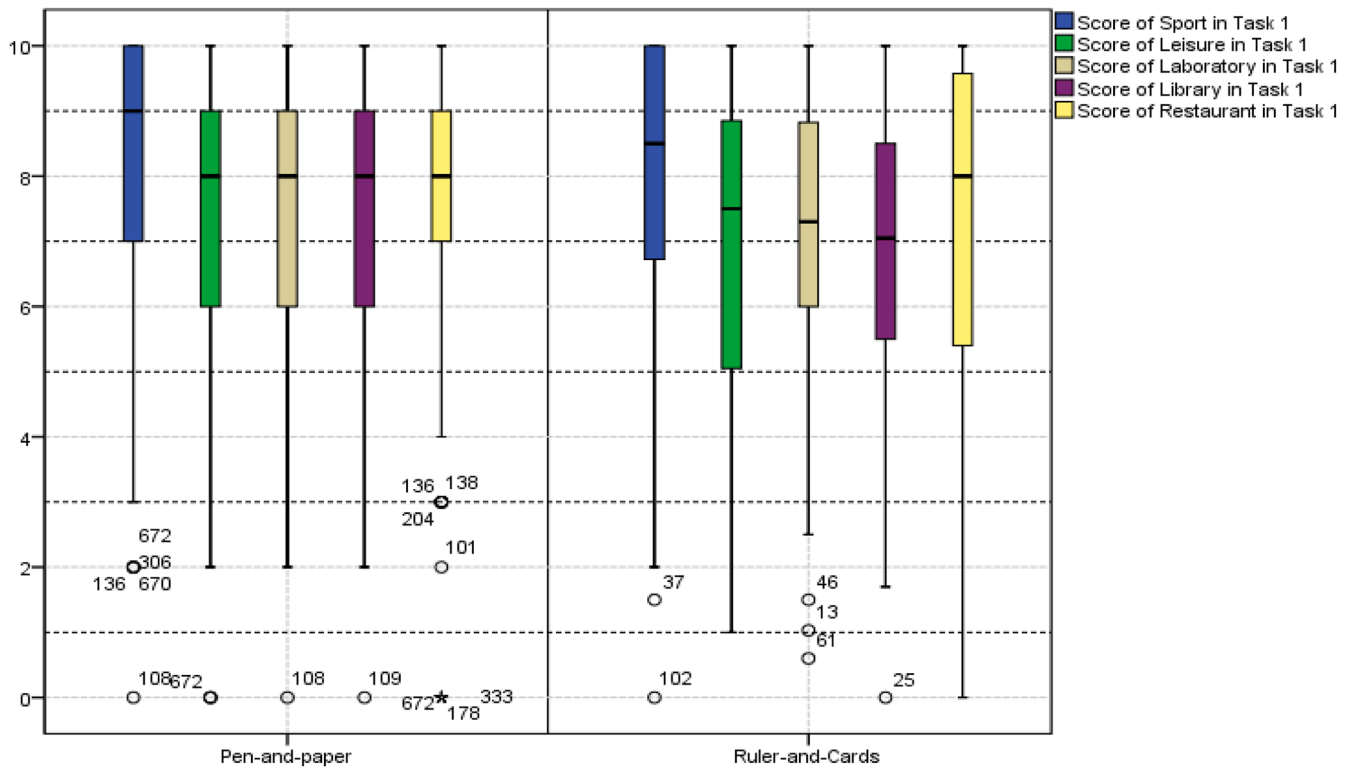


Fig. 5. Ranking of resource attributes scores for pen-and-paper and ruler-and-cards.

interviewers employed the ruler-and-cards treatment ($P_{5,2}$). The null hypothesis is that there is no difference between the median ranks of all interviewers regarding consistency levels. We reject the null hypothesis ($H(2) = 5.143, p = .050$). Different interviewers may be associated with variations in consistency levels.

Proposition 6. For the assisted mode, the application sequence (ruler-and-cards versus cards-and-ruler) influences:

$P_{6,1}$. The transitivity in the resource attribute ranking in the proposed protocol.

$P_{6,2}$. The consistency between the SW method for survey-based preference elicitation and the ordinal ranking of preferences elicited from the proposed protocol.

The χ^2 test was conducted to check whether there is a difference in rank between the sequences of two groups (ruler-and-cards or cards-and-ruler) for the transitivity in the resource attribute ranking in the proposed protocol ($P_{6,1}$) (null hypothesis). We obtained the null hypothesis $H(2) = 2556$ and $p\text{-value} = 0.110$. We conclude that the sequences of tasks (R-C versus C-R) could not be significantly associated with the observed transitivity.

For $P_{6,2}$, the Mann-Whitney U test was employed to determine whether there is a difference in rank between the sequence of two groups (ruler-and-cards or cards-and-ruler) according to the level of consistency in the judgments. The results indicated a non-significant difference between the two groups ($U = 1358,500, p = .215$ for consistency levels). Additionally, the sequences of tasks (R-C versus C-R) were not significantly associated with the consistency levels.

7. Discussion

In this section, we analyse and discuss the results of the behavioural experiment. Sections 7.1 and 7.2 address the first and second research questions, respectively.

7.1. Is the swing weighting method suitable for survey-based preference elicitation in resource allocation decisions?

7.1.1. Process compliance in the survey-based SW procedure (assisted and unassisted modes)

In our behavioural experiment, we identified three possible issues regarding process compliance in the survey-based SW method. First, many subjects did not follow the first basic instruction for the pen-and-paper treatment: “to rate 10 points to the most preferred hypothetical alternative (i.e., first swing in the resource moving from the status quo to the ideal level)”. This was the case even though the completion time of the questionnaire was only about 10 min, and its format was carefully designed to avoid fatigue or discomfort for the subjects to answer it. This issue has been reported in previous studies, e.g. by Aubert et al. (2020), who found process compliance failures in their intervention (only 4.7 % followed all the instructions of the process); similar results were found by Haag et al. (2022); roughly 88 % of the responses did not fully comply with the swing weighting instructions.

Second, some subjects did not perform the second instruction in the pen-and-paper mode: “to complete the subsequent swings of resources, comparing it with the first swing”. Past studies highlighted similar issues, with the scoring part of the swing method being reported as challenging even when using an interactive mode (Aubert et al., 2022) and time-consuming (Aubert et al., 2020).

Third, the subjects did not use non-integer value scores on the continuum scale if they were in the pen-and-paper mode. One possible explanation is that the subjects provided direct ratings, as they are accustomed to filling out marketing surveys, as Aubert et al. (2020) pointed out. The subjects may have provided simple importance levels or beliefs about resource attributes (Rosenberg, 1956; Fishbein, 1963; Bass & Talarzyk, 1972), treating the swing weights as if they were absolute (a priori) weights not tied to the first swing (Danielson & Ekenberg, 2019). On the other hand, in the ruler treatment (side without scale), the subjects did use non-integer scores. This finding is in line with a study conducted by Pöyhönen and Hämäläinen (2001), in which 62 %

(of 247) subjects scored multiples of tens.

Our conclusion for this compliance issue in employing the SW method for survey-based elicitations in our behavioural experiment is that the subjects had found the tasks required by the pen-and-paper SW elicitation protocol as challenging. There is also some indicative evidence that their judgments were not necessarily following value-distance assessments against the first swing weight, as required by the method. Hence, the survey-based SW may need to be conducted in an assisted mode with a protocol that encourages correct reasoning to form these value judgments, instead of a standard SW protocol as currently adopted in many DA community-based interventions.

7.1.2. (In) consistency of preferences for resource attributes between tasks

Our results showed low individual consistency levels between the two tasks (SW versus the ordinal ranking in our protocol) under both treatments. Indeed, even in the assisted mode, many subjects had inconsistent ranks between the two tasks. Furthermore, if we consider the overall consistency level for only the subjects with transitive preferences, their consistency level between tasks were slightly higher for the ruler-and-cards treatment than for the pen-and-paper treatment. This finding shows that the survey-based SW often fails the procedural invariance test, which requires that weights elicited by different tasks (methods) should be the same if the subjects consistently assess them (Mustajoki et al., 2005).

There are some possible additional causes for this inconsistency between rankings from the two tasks, beyond the possible complexity of survey-based SW, such as the lack of familiarity with the methods (Park & Lessig, 1981) and the number of resource attribute levels considered (Borcherding et al., 1991; Pöyhönen & Hämäläinen, 2001). These findings and those from Section 7.1.1 suggest the SW method may not be advised to measure the value of resource attributes in a survey-based mode (van Ittersum et al., 2007). In addition, it indicates that using the SW method in this mode would require additional consistency checks of preferences (Zardari et al., 2015), replicating best Decision Analysis practices in which the decision analyst conducts these checks (von Winterfeldt & Edwards, 1986).

7.1.3. Consistency of first and last resource attribute ranking between tasks

The experimental results showed low consistency of first and last resource attribute ranking between the two tasks (SW versus the ordinal ranking in our protocol). While this inconsistency could have been explained if a large number of attributes were utilised (Pöyhönen & Hämäläinen, 2001), our experimental design employed only five resource attributes. It is also plausible that an abstract nature and the relative detachment of the resource attribute could have prevented the subjects from internalising the swings. However, we designed the experiment to make the decision setting as realistic as possible, reducing such risk.

The other possible cause for this inconsistency is when the value scores rated by the subjects were close to one another. Considering the ruler-and-cards treatment, 83 % scored close values for resource attributes in Task 1 (SW method) regarding the subjects who were inconsistent between tasks. The definition of a threshold for partial inconsistency (see the second note in Table 2) tried to address this issue in the data analysis.

We also analysed the consistency of the resource attribute ranking among tasks only for those subjects who revealed transitivity in their ranking of resource attributes in Task 2. In this specific case, we carried out an additional analysis considering only subjects with transitive preferences. The results show that 24% (95% CI [18.39%, 29.95%]) of the subjects selected different first resource attributes and 27% (95% CI [20.58%, 32.50%]) selected different last resource attributes among tasks for both treatments. Even in the assisted mode, 21% (95% CI [13.16%, 29.27%]) and 27% (95% CI [18.50%, 36.05%]) of the subjects selected different resource attributes (first and last, respectively).

This experimental finding is a major concern for researchers who

intend to employ the SW method in a survey-based mode. The method relies heavily on the adequate identification of the first resource attribute, which is then used to evaluate the relative value of the subsequent resource attributes. If a first resource attribute ranking cannot be adequately identified and consistently maintained, the relative valuations of the resource attributes would be seriously compromised.

7.1.4. Range in the SW procedure

The ruler-and-cards treatment was an assisted mode in which the survey interviewer conducted the task throughout the elicitation steps. In this treatment, subjects used the continuum scale by moving the resource swings along the ruler and using crayons to rate the weights among resource attributes; therefore, there is evidence that the subjects indicated the distance on the ruler while they revealed their preferences. Such assessment is important because the subjects could prioritise the dimensions consistently, as they were supported step-by-step when they needed to evaluate the next resource attribute and compare it to the most valuable swing already chosen. Therefore, the ruler-based mode offered a wider range of valuations than the paper-based mode, with more spread in the subjects' responses (a larger interquartile range size).

In the pen-and-paper treatment, the subjects may have directly scored the resource attributes instead of considering their relative value, as previously alluded. Since the subjects did not use non-integer value scores and most revealed indifference (by using the same value scores among resource attributes), this treatment may fail to meet the range sensitivity requirement (Fischer, 1995). The response scale effects may lead subjects to ignore ranges (Borcherding et al., 1991) or the elicited weights may represent general values or attitudes toward the criteria, not specific trade-offs among them (Stewart & Ely, 1984).

Our conclusion when analysing the ranges in the SW method is that responses from this method benefit from a survey researcher who elicits the preferences of subjects, ensuring the possibility of employing a wider spread of valuations. However, this assisted mode is unfeasible for a large sample size. Conversely, pen-and-paper may elicit more homogeneous responses with a smaller spread of valuations. Partially, this may be a consequence of the 0 to 10 scale that we employed (which matches the university evaluation score of the students). Still, we conjecture that this lack of spread might result from subjects attributing 'marks' to resources instead of valuing the resource attribute swings.

7.2. To what extent can the proposed protocol provide coherent responses in resource allocation decisions?

To implement our elicitation protocol, we evaluated five resource attributes in pairs of profiles, which resulted in ten pairwise comparisons in Task 2. There was a high transitivity in Task 2, regardless of the treatment. Therefore, our proposed protocol can be applied either as a survey or by survey interviewers. The choice of application mode will depend on the subjects, as the assisted mode can be simpler for subjects with less formal education levels. Furthermore, while different survey interviewers may influence the consistency level among subjects, these interviewers have not influenced the transitivity of the subjects. The suggested protocol requires strict preferences, and most subjects in the behavioural experiment expressed transitive preferences.

However, most of the subjects who were transitive (211) in Task 2 were inconsistent between tasks (123/211 or 58 % of the sample), with roughly the same proportion for both treatments (68 % for pen-and-paper treatment and 70 % for ruler-and-cards treatment). There was no difference in resource attribute preferences in the application sequence of treatments (ruler-and-cards or cards-and-ruler). Hence, the SW method could be employed in a survey only for those subjects who maintained transitive preferences in a sequence of cards followed by the ruler.

On the other hand, our proposed protocol replicates the ranking from SWs for those subjects that have expressed transitive preferences. Encouragingly, even the subjects who did not comply with the SW

procedure performed pairwise comparisons with relatively high transitivity in our protocol (82.6 % = 19/23 of the subjects). While intransitive preferences (17.4 % of the sample) are concerning, their identification may help exclude their responses from the sample or re-elicit the ranking with the support of a survey interviewer.

Typically, decision-makers show resistance against sophisticated elicitation methods (Aloysius et al., 2006; Kottemann & Davis, 1991). They prefer more straightforward and easier-to-understand methods, such as pairwise comparisons (Zheng & Lienert, 2018), which are also perceived to be more accurate (Aloysius et al., 2006). Our protocol has shown a high transitivity of preferences in the behavioural experiment (with assisted or unassisted support). Hence, the protocol has performed consistently well for both treatments (pen-and-paper and ruler-and-cards). In addition, the protocol is straightforward, with clear questions and profile options, and no process compliance failures occurred (even in the pen-and-paper format). Judging strict preference between two profiles is cognitively easier and less prone to noise in the elicitation (both in the assisted and unassisted modes) than assessing the value assessment of dummy alternatives required SWs (see also Larichev (1992)).

8. Conclusion and directions for further research

The efficient allocation of scarce resources to improve community services and facilities is a key aspect of Community OR. Decision Analysis (DA) is being increasingly employed to support such resource allocation processes, as it enables the evaluation of options on competing objectives and has the potential to represent the priorities of a community. In addition, current attempts to increase the representativeness of the community in Decision Analysis interventions have led to the use of survey-based preference elicitation, typically adapting elicitation methods, such as the Swing Weighting (SW) method, to (online) surveys. However, some previous studies have indicated that subjects experienced challenges in having their preferences elicited via this mode, which may have a potential impact on the coherence of the preferences being elicited.

In this paper, we suggested a survey-based priority elicitation protocol for communities that can be applied on a large scale while maintaining the rigour required by DA. The proposed protocol is straightforward and requires limited cognitive effort, eliciting only strict preference relations. The protocol aims to address these requirements to be applied for any community, particularly deprived ones, in which subjects have low levels of formal education. We argue that strict preferences are suitable for resource allocation decisions, as clear priorities must be expressed at an individual level; thus, the protocol does not allow indifference statements.

We tested the protocol in a hypothetical (but realistic) resource allocation problem: a controlled behavioural experiment in which subjects were asked to prioritise attribute resources. Two treatments were developed: an unassisted mode (pen-and-paper) and an assisted mode (ruler-and-cards), both composed of the SW method (Task 1) and the proposed protocol (Task 2). We also compared, in our behavioural experiment, the results obtained from the suggested protocol with responses from the survey-based SW method and found a relatively large number of violations in the latter method (e.g. rank reversals against the ordinal ranking, inconsistent use of first resource attribute ranking, some indications that value difference judgments were not being made).

We thus conclude that the standard SW method may not be suitable for survey-based preference elicitation in resource allocation problems, due to the relative complexity of the elicitation protocol and some inconsistent results against ordinal preferences that we found in our behavioural experiment. However, more behavioural evidence would be welcome to confirm this finding and extend to other types of multi-attribute prioritisation, as well as assess the suitability of employing other standard DA preference elicitation methods for attribute weighting in online surveys (e.g., the trade-off method).

Encouragingly, our experimental findings suggest that our proposed protocol could be used to screen out subjects with intransitive preferences, before the SW method is employed in surveys. This screening step could potentially increase the coherence of the elicited preferences from this widely employed priority elicitation method.

We can identify some relevant avenues for further research, which are briefly highlighted next. First, the development of a consistency index for the ordinal data elicited from individual subjects would be welcome. Such an index could borrow ideas from the indices suggested by Liang et al. (2022) or the proposed inconsistencies index between statements using scale-independent consistency measures (see Salo and Hämäläinen, 1995; Salo, 1993). In addition, the card method could be adapted to prioritise disinvestment decisions, given the current pressures on public budgets.

Second, a cognitive assessment of reasoning efforts and cognitive style could shed light on the subject's reasoning when answering the survey. For example, the cognitive reflection test (Thomson & Oppenheimer, 2016) could be applied to examine which cognitive aspects affect their answers. A comparison with the Deck of Cards method (Corrente et al., 2021; Figueira & Roy, 2002; Siskos & Tsotsolas, 2015) could also shed some light on the level of reasoning effort demanded by the proposed elicitation protocol versus the type of preference information that is provided by the subjects (Corrente et al., 2021). Another interesting comparison is whether the mode of survey application for the proposed protocol, either with pen-and-paper or with cards supported by a survey interviewer, would generate different rankings.

Third, our results are based on a behavioural experiment which, albeit adopting a realistic decision context, is somehow artificial. This experimental setting is prevalent in Behavioural Decision Analysis (Federspiel et al., 2024), maximising experimental control over realism (McGrath, 1981). Nonetheless, field tests of the protocol in more realistic settings could improve our understanding of its benefits and drawbacks.

Fourth, it is important to investigate how to derive community attribute weights after eliciting individual rankings, given the intrinsic challenge created by Arrow's paradox (Arrow, 1950) when only ordinal individual rankings are considered (French, 2007). As the protocol has been conceptualised within a MAVT framework, such attribute weights are scaling constants representing value trade-offs. We distinguish approaches to derive these weights into three categories:

- Translate the individual ordinal rank into individual quantitative weights and aggregate these individual weights at the community level (individual weights) (see rank-sum (Barron & Barrett, 1996); rank reciprocal (Stillwell & Seaver, 1981); rank-order centroid (Edwards & Barron, 1994); and combination of the Centroid (Olson & Dorai, 1992)).
- Employ the individual ordinal rank information and identify quantitative weights at the community level that are compatible with the individual ordinal rank (community weights) (e.g. by employing the simulation technique provided by Butler et al. (1997)).
- Infer community-level priorities by employing preference learning algorithms (Chevaleyre et al., 2010) or judgment aggregation for ordinal preferences (Grossi & Pigozzi, 2014; Wilson, 1975), as well as considering coalitions with similar weights among groups within a community.

The first approach obfuscates individual rankings. They are also rank-based methods highly sensitive to the number of resource attributes and may provide rather arbitrary quantitative weights. On the other hand, in the second approach, the exact weights from an assessment procedure may be legitimately questioned, given the transition from individual ordinal ranks to the community-level quantitative weights identified by the Monte-Carlo simulation. Thus, it remains an open question what the best way of aggregating the individual rankings is.

Concluding this paper, we find survey-based preference elicitation an exciting area for research in DA and Community OR, given the increasing relevance of online surveys and contemporary attempts to make resource allocation more inclusive and democratic. We hope that this paper can help the OR community to further develop adequate survey-based preference elicitation methods for community-based resource allocation processes.

CRedit authorship contribution statement

José Geraldo Vidal Vieira: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Gilberto Montibeller:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Investigation, Conceptualization.

Acknowledgements

We are grateful to the three anonymous reviewers for the thoughtful comments that have helped to significantly improve our draft, and to the editor Mike Yearworth for the opportunity to revise and further develop the paper. We thank the three survey interviewers Kimberlly Cury Barros Christiano, Livia Maria Marchi, Paola Viviana Campos Tinoco for their support in the collection of the data. We also thank Luis Alberto Araújo Dorado for the administrative support in the preparation of the supplementary material. Funding: This work was supported by the FAPESP Brazilian agency (Grant# 202202338-8).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ejor.2025.07.049](https://doi.org/10.1016/j.ejor.2025.07.049).

References

- Aloysius, J. A., Davis, F. D., Wilson, D. D., Taylor, A. R., & Kottemann, J. E. (2006). User acceptance of multi-criteria decision support systems: The impact of preference elicitation techniques. *European Journal of Operational Research*, 169(1), 273–285. <https://doi.org/10.1016/j.ejor.2004.05.031>
- Aribarg, A., Burson, K. A., & Larrick, R. P. (2017). Tipping the scale: The role of discriminability in conjoint analysis. *Journal of Marketing Research*, 54(2), 279–292. <https://doi.org/10.1509/jmr.14.0659>
- Arrow, K. J. (1950). A difficulty in the concept of social welfare. *The Journal of Political Economy*, 58(4). <https://doi.org/10.1086/256963>. Issue.
- Aubert, A. H., Esculier, F., & Lienert, J. (2020). Recommendations for online elicitation of swing weights from citizens in environmental decision-making. *Operations Research Perspectives*, 7. <https://doi.org/10.1016/j.orp.2020.100156>
- Aubert, A. H., & Lienert, J. (2019). Gamified online survey to elicit citizens' preferences and enhance learning for environmental decisions. *Environmental Modelling and Software*, 111, 1–12. <https://doi.org/10.1016/j.envsoft.2018.09.013>
- Aubert, A. H., Schmid, S., Beutler, P., & Lienert, J. (2022). Innovative online survey about sustainable wastewater management: What young Swiss citizens know and value. *Environmental Science and Policy*, 137, 323–335. <https://doi.org/10.1016/j.envsci.2022.08.018>
- Aubert, A. H., Lienert, J., & von Helversen, B. (2023). Gamified environmental multi-criteria decision analysis: Information on objectives and range insensitivity bias. *International Transactions in Operational Research*, 30(6), 3738–3770. <https://doi.org/10.1111/itor.13206>
- Aubert, A. H., Schmid, S., & Lienert, J. (2024). Can online interfaces enhance learning for public decision-making? Eliciting citizens' preferences for multicriteria decision analysis. *European Journal of Operational Research*, 314(2), 760–775. <https://doi.org/10.1016/j.ejor.2023.10.031>
- Bana E Costa, C. A. (2001). The use of multi-criteria decision analysis to support the search for less conflicting policy options in a multi-actor context: Case study. *Journal of Multi-Criteria Decision Analysis*, 10(2), 111–125. <https://doi.org/10.1002/mcda.292>
- Barron, F. H., & Barrett, B. E. (1996). Decision quality using ranked attribute weights. *Management Science*, 42(11), 1515–1523. <https://doi.org/10.1287/mnsc.42.11.1515>
- Bass, F. M., & Talarzyk, W. W. (1972). An attitude model for the study of brand preference. *Journal of Marketing Research*, 9, 92–96. <https://doi.org/10.1177/002224377200900121>
- Batsell, R. R. (1980). Consumer resource allocation models at the individual level. *Source: Journal of Consumer Research*, 7(1), 78–87. <https://doi.org/10.1086/208795>
- Belton, V., & Stewart, T. J. (2002). *Multiple criteria decision analysis: An integrated approach*. Dordrecht: Kluwer: Kluwer. <https://doi.org/10.1007/978-1-4615-1495-4>
- Borcherding, K., Eppel, T., & von Winterfeldt, D. (1991). Comparison of weighting judgments in multiattribute utility measurement. *Management Science*, 37(12). <https://doi.org/10.1287/mnsc.37.12.1603>. Issue.
- Bryson, N., Mobolurin, A., & Ngwenyama, O. (1995). Modelling pairwise comparisons on ratio scales. *European Journal of Operational Research*, 83, 639–654. <https://doi.org/10.1016/0.2.1.4793.9366>
- Butler, J., Jia, J., & Dyer, J. (1997). Simulation techniques for the sensitivity analysis of multi-criteria decision models. *European Journal of Operational Research*, 103(3), 531–546. [https://doi.org/10.1016/S0377-2217\(96\)00307-4](https://doi.org/10.1016/S0377-2217(96)00307-4)
- Carson, R. T., Louviere, J. J., Anderson, D. A., Arabe, P., Bunch, D. S., Hensher, D. A., Johnson, R. M., Kuhfeld, W. F., Steinberg, D., Swait, J., Timmermans, H., & Wiley, J. B. (1994). Experimental analysis of choice. *Marketing Letters*, 5(4), 351–368. <https://doi.org/10.1007/BF00999210>
- Chevalere, Y., Koriche, F., Lang, J., Mengin, J., Zanuttini, B., & Learning, B. Z. (2010). Ordinal preferences on multiattribute domains: The case of CP-nets. In J. Fürnkranz, & E. Hüllermeier (Eds.), *Preference learning* (pp. 273–296). Springer. https://doi.org/10.1007/978-3-642-14125-6_131
- Corrente, S., Figueira, J. R., & Greco, S. (2021). Pairwise comparison tables within the deck of cards method in multiple criteria decision aiding. *European Journal of Operational Research*, 291(2), 738–756. <https://doi.org/10.1016/j.ejor.2020.09.036>
- Danielson, M., & Ekenberg, L. (2019). An improvement to swing techniques for elicitation in MCDM methods. *Knowledge-Based Systems*, 168, 70–79. <https://doi.org/10.1016/j.knsys.2019.01.001>
- Eden, C. (1992). *A framework for thinking about group decision support systems (GDSS)*, 1. Group Decision and Negotiation. <https://doi.org/10.1007/BF00126263>
- Edwards, W., & Barron, F. H. (1994). SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes*, 60(3), 306–325. <https://doi.org/10.1093/acprof:oso/9780195322989.003.0031>
- Falk, A., Becker, A., Dohmen, T., Huffman, D., & Sunde, U. (2022). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science*. <https://doi.org/10.1287/mnsc.2022.4455>
- Figueira, J., & Roy, B. (2002). Decision aiding determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. *European Journal of Operational Research*, 139, 317–326. [https://doi.org/10.1016/S0377-2217\(01\)00370-8](https://doi.org/10.1016/S0377-2217(01)00370-8)
- Fischer, G. (1995). Range sensitivity of attribute weights in multiattribute value. *Organizational Behavior and Human Decision Processes*, 62(3), 252–266. <https://doi.org/10.1006/obhd.1995.1048>
- Fishbein, M. (1963). An investigation of the relationships between beliefs about an object and the attitude toward that object. *Human Relations*, 16, 233–240. <https://doi.org/10.1177/001872676301600302>
- Franco, L. A., & Montibeller, G. (2010). Facilitated modelling in operational research. *European Journal of Operational Research*, 205(3), 489–500. <https://doi.org/10.1016/j.ejor.2009.09.030>
- Federspiel, F., Montibeller, G., & Seifert, M. (2024). Behavioral decision analysis: Past, present and future. In F. Federspiel, G. Montibeller, & M. Seifert (Eds.), *Behavioral decision analysis: 350. Behavioral decision analysis*. Springer. https://doi.org/10.1007/978-3-031-44424-1_1
- French, S. (2007). Web-enabled strategic GDSS, e-democracy and Arrow's theorem: A Bayesian perspective. *Decision Support Systems*, 43(4), 1476–1484. <https://doi.org/10.1016/j.dss.2006.06.003>
- Fürnkranz, J., & Hüllermeier, E. (2010). *Preference Learning*. <https://doi.org/10.1007/978-3-642-14125-6>
- Green, P., & Wind, Y. (1975). New way to measure consumers' judgments. *Harvard Business Review*, 53, 107–117.
- Grossi, D., & Pigozzi, G. (2014). In D. Grossi, & D. Grossi (Eds.), *Judgment aggregation: A primer*. Cham: Springer. <https://doi.org/10.1007/978-3-031-01568-7>
- Haag, F., Aubert, A. H., & Lienert, J. (2022). ValueDecisions, a web app to support decisions with conflicting objectives, multiple stakeholders, and uncertainty. *Environmental Modelling and Software*, 150. <https://doi.org/10.1016/j.envsoft.2022.105361>
- Haag, F., Zürcher, S., & Lienert, J. (2019). Enhancing the elicitation of diverse decision objectives for public planning. *European Journal of Operational Research*, 279(3), 912–928. <https://doi.org/10.1016/j.ejor.2019.06.002>
- Huber, J., Wittink, D., Fielder, J., & Miller, R. (1993). The effectiveness of alternative preference elicitation procedures in predicting choice. *Journal of Marketing Research*, 30, 105–114. <https://doi.org/10.1177/002224379303000109>
- Hummel, J. M., Oliveira, M. D., Bana e Costa, C. A., & IJerman, M. J. (2017). Supporting the project portfolio selection decision of research and development investments by means of multi-criteria resource allocation modelling. *Multi-Criteria decision analysis to support healthcare decisions* (pp. 89–103). Springer International Publishing. https://doi.org/10.1007/978-3-319-47540-0_6
- Keeney, R. L. (1988). Building models of values. *European Journal of Operational Research*, 37(2), 149–157. [https://doi.org/10.1016/0377-2217\(88\)90324-4](https://doi.org/10.1016/0377-2217(88)90324-4)
- Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences end value trade-offs*. Cambridge university press. <https://doi.org/10.1017/CBO9781139174084.011>
- Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value trade-offs*. New York: John Wiley. <https://doi.org/10.1017/CBO9781139174084>
- Kimbrough, S. O., & Weber, M. (1994). An empirical comparison of utility assessment programs. *European Journal of Operational Research*, 75(3), 617–633. [https://doi.org/10.1016/0377-2217\(94\)90301-8](https://doi.org/10.1016/0377-2217(94)90301-8)

- Kottemann, J. E., & Davis, F. D. (1991). Decisional conflict and user acceptance of multicriteria decision-making aids*. *Decision Sciences*, 22(4), 918–927. <https://doi.org/10.1111/j.1540-5915.1991.tb00371.x>
- Kuller, M., Beutler, P., & Lienert, J. (2022). Preference change in stakeholder group-decision processes in the public sector: Extent, causes and implications. *European Journal of Operational Research*. <https://doi.org/10.1016/j.ejor.2022.12.001>
- Larichev, I. (1992). Cognitive validity in design of decision-aiding techniques. *Journal of Multi-Criteria Decision Analysis*, 1, 127–138. https://doi.org/10.1007/978-3-642-57311-8_30
- Larichev, O. I., & Moshkovich, H. M. (1995). ZAPROS-LM-A method and system for ordering multiattribute alternatives. *European Journal of Operational Research*, 82, 503–521. [https://doi.org/10.1016/0377-2217\(93\)e0143-1](https://doi.org/10.1016/0377-2217(93)e0143-1)
- Liang, F., Brunelli, M., & Rezaei, J. (2022). Best-worst tradeoff method. *Information Sciences*, 610, 957–976. <https://doi.org/10.1016/j.ins.2022.07.097>
- Lienert, J., Duygan, M., & Zheng, J. (2016). Preference stability over time with multiple elicitation methods to support wastewater infrastructure decision-making. *European Journal of Op.Res.*, 253(3), 746–760. <https://doi.org/10.1016/j.ejor.2016.03.010>
- Liesjö, J., Andelmin, J., & Salo, A. (2020). Efficient allocation of resources to a portfolio of decision making units. *European Journal of Operational Research*, 286(2), 619–636. <https://doi.org/10.1016/j.ejor.2020.03.031>
- Liesjö, J., Mild, P., & Salo, A. (2007). Preference programming for robust portfolio modeling and project selection. *European Journal of Operational Research*, 181(3), 1488–1505. <https://doi.org/10.1016/j.ejor.2005.12.041>
- Louviere, J. J. (1988). *Analyzing decision making: Metric conjoint analysis*, 67. Sage. <https://doi.org/10.1177/014662169001400112>. Vol.
- Luce, D. (1959). *Individual choice behavior: A theoretical analysis*. Dover Publication.
- McGrath, J. E. (1981). Dilemmas: The study of research choices and dilemmas. *American Behavioral Scientist*, 25(2), 179–210. <https://doi.org/10.1177/000276428102500205>
- Montibeller, G. (2018). Behavioral challenges in policy analysis with conflicting objectives. *Recent advances in optimization and modeling of contemporary problems* (pp. 85–108). INFORMS. <https://doi.org/10.1287/educ.2018.0182>
- Montibeller, G., & von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk Analysis*, 35(7), 1230–1251. <https://doi.org/10.1111/risa.12360>
- Montibeller, G., Franco, L. A., Lord, E., & Iglesias, A. (2009). Structuring resource allocation decisions: A framework for building multi-criteria portfolio models with area-grouped options. *European Journal of Operational Research*, 199(3), 846–856. <https://doi.org/10.1016/j.ejor.2009.01.054>
- Moshkovich, H. M., Mechtov, A. I., & Olson, D. L. (2002). Ordinal judgments in multiattribute decision analysis. *European Journal of Operational Research*, 137(3), 625–641. [https://doi.org/10.1016/S0377-2217\(01\)00106-0](https://doi.org/10.1016/S0377-2217(01)00106-0)
- Mustajoki, J., Hämäläinen, R. P., & Salo, A. (2005). Decision support by interval SMART/SWING - incorporating imprecision in the SMART and SWING methods. *Decision Sciences*, 36(2), 317–339. <https://doi.org/10.1111/j.1540-5414.2005.00075.x>
- Ngwenyama, O., & Bryson, N. (1999). Eliciting and mapping qualitative preferences to numeric rankings in group decision making. *European Journal of Operational Research*, 116, 487–497. [https://doi.org/10.1016/S0377-2217\(98\)00081-2](https://doi.org/10.1016/S0377-2217(98)00081-2)
- Olson, D. L., & Dorai, V. K. (1992). Implementation of the centroid method of Solymosi and Dombi. *European Journal of Operational Research*, 60. [https://doi.org/10.1016/0377-2217\(92\)90339-B](https://doi.org/10.1016/0377-2217(92)90339-B)
- Park, C. W., & Lessig, V. P. (1981). Familiarity and its impact on consumer decision biases and heuristics. *Journal of Consumer Research*, 8(2), 223–230. <https://doi.org/10.1086/208859>
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral Decision Research: A constructive processing perspective. *Annual Review of Psychology*, 43, 87–131. <https://doi.org/10.1146/annurev.ps.43.020192.000511>
- Phillips, L. D., & Bana e Costa, C. A. (2007). Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Annals of Operations Research*, 154(1), 51–68. <https://doi.org/10.1007/s10479-007-0183-3>
- Phillips, L. D. (2007). In W. Edwards, R. Miles Jr, & D. von Winterfeldt (Eds.), *Decision conferencing. in advances in decision analysis from foundations to applications* (pp. 375–399). Cambridge University Press. ISBN 9780521682305.
- Pictet, J., & Bollinger, D. (2008). Extended use of the cards procedure as a simple elicitation technique for MAVT. Application to public procurement in Switzerland. *European Journal of Operational Research*, 185(3), 1300–1307. <https://doi.org/10.1016/j.ejor.2006.05.051>
- Pöyhönen, M., & Hämäläinen, R. P. (2001). On the convergence of multiattribute weighting methods. *European Journal of Operational Research*, 129(3), 569–585. [https://doi.org/10.1016/S0377-2217\(99\)00467-1](https://doi.org/10.1016/S0377-2217(99)00467-1)
- Rao, V. R. (2014). *Applied conjoint analysis*. Springer. <https://doi.org/10.1007/978-3-540-87753-0>
- Rezaei, J. (2021). Anchoring bias in eliciting attribute weights and values in multi-attribute decision-making. *Journal of Decision Systems*, 30(1), 72–96. <https://doi.org/10.1080/12460125.2020.1840705>
- Rezaei, J., Arab, A., & Mehregan, M. (2022). Analyzing anchoring bias in attribute weight elicitation of SMART, swing, and best-worst method. *International Transactions in Operational Research*. <https://doi.org/10.1111/itor.13171>
- Riabacke, M., Danielson, M., & Ekenberg, L. (2012). State-of-the-art prescriptive criteria weight elicitation. In *Advances in decision sciences, 2012*. <https://doi.org/10.1155/2012/276584>
- Rosenberg, M. J. (1956). Cognitive structure and attitudinal affect. *Journal of Abnormal and Social Psychology*, 53, 367–372. <https://doi.org/10.1037/h0044579>
- Salo, A. A. (1993). Inconsistency analysis by approximately specified priorities. *Mathematical and Computer Modelling*, 17(4–5), 123–133. [https://doi.org/10.1016/0895-7177\(93\)90181-W](https://doi.org/10.1016/0895-7177(93)90181-W)
- Salo, A. A., & Hämäläinen, R. P. (1995). Preference programming through approximate ratio comparisons. *European Journal of Operational Research*, 82(3), 458–475. [https://doi.org/10.1016/0377-2217\(93\)E0224-L](https://doi.org/10.1016/0377-2217(93)E0224-L)
- Salo, A., Keisler, J., & Morton, A. (2011). *Portfolio decision analysis: Improved methods for resource allocation*. Springer. <https://doi.org/10.1007/978-1-4419-9943-6>
- Schlereth, C., Eckert, C., Schaaf, R., & Skiera, B. (2014). Measurement of preferences with self-explicated approaches: A classification and merge of trade-off- and non-trade-off-based evaluation types. *European Journal of Operational Research*, 238(1), 185–198. <https://doi.org/10.1016/j.ejor.2014.03.010>
- Scholten, L., Schuwirth, N., Reichert, P., & Lienert, J. (2015). Tackling uncertainty in multi-criteria decision analysis – An application to water supply infrastructure planning. *European Journal of Operational Research*, 242(1), 243–260. <https://doi.org/10.1016/j.ejor.2014.09.044>
- Scholz, S., Meissner, M., & Decker, R. (2010). Measuring consumer preferences for complex products: A compositional approach based on paired comparisons. *Journal of Marketing Research*, 47(4), 685–698. <https://doi.org/10.1509/jmkr.47.4.685>
- Siskos, E., & Tsotsolas, N. (2015). Elicitation of criteria importance weights through the Simos method: A robustness concern. *European Journal of Operational Research*, 246(2), 543–553. <https://doi.org/10.1016/j.ejor.2015.04.037>
- Srinivasan, V. (1988). A conjunctive-compensatory approach to the self-explication of multiattributed preferences. *Decision Sciences*, 19(2), 295–305. <https://doi.org/10.1111/j.1540-5915.1988.tb00268.x>
- Stewart, T. R., & Ely, D. W. (1984). Range sensitivity: A necessary condition and a test for the validity of weights. *Unpublished manuscript*. Boulder, CO: National Center for Atmospheric Research. https://www.researchgate.net/publication/237537318_Range_sensitivity_A_necessary_condition_and_a_test_for_the_validity_of_weights
- Stillwell, W. G., & Seaver, D. A. (1981). A comparison of weight approximation techniques in multiattribute utility decision making. *Organizational behavior and human performance*, 28(1), 62–77. [https://doi.org/10.1016/0030-5073\(81\)90015-5](https://doi.org/10.1016/0030-5073(81)90015-5). Vol.
- Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, 11(1), 99–113. <https://doi.org/10.1017/s1930297500007622>
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological review*, 79(4), 281–299. <https://doi.org/10.1037/h0032955>
- van Ittersum, K., Pennings, J. M. E., Wansink, B., & van Trijp, H. C. M. (2007). The validity of attribute-importance measurement: A review. *Journal of Business Research*, 60(11), 1177–1190. <https://doi.org/10.1016/j.jbusres.2007.04.001>
- von Nitzsch, R., & Weber, M. (1993). *The effect of attribute ranges on weights in multiattribute utility measurements*. <https://doi.org/10.1287/mnsc.39.8.937>
- von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge University Press. <https://doi.org/10.1177/0272989x8700700312>
- Weber, M., & Borchering, K. (1993). Behavioral influences on weight judgments in multiattribute decision making. *European Journal of Operational Research*, 67, 1–12. [https://doi.org/10.1016/0377-2217\(93\)90318-H](https://doi.org/10.1016/0377-2217(93)90318-H)
- Wilson, R. (1975). On the theory of aggregation. *Journal of economic theory*, 10, 89–99. [https://doi.org/10.1016/0022-0531\(75\)90062-9](https://doi.org/10.1016/0022-0531(75)90062-9)
- Zardari, N. H., Ahmed, K., Shirazi, S. M., & Yusop, Z. B. (2015). *Weighting methods and their effects on multi-criteria decision making model outcomes in water resources management*. <https://doi.org/10.1007/978-3-319-12586-2>. <http://www.springer.com/series/11214>
- Zheng, J., & Lienert, J. (2018). Stakeholder interviews with two MAVT preference elicitation philosophies in a Swiss water infrastructure decision: Aggregation using SWING-weighting and disaggregation using UTAGMS. *European Journal of Operational Research*, 267(1), 273–287. <https://doi.org/10.1016/j.ejor.2017.11.018>