The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1535-0118.htm

JOPP
22,2

Received 16 June 2021
Revised 6 October 2021
Accepted 6 February 2022

Choosing the most economically advantageous tender using a multi-criteria decision analysis approach

Juha-Matti Lehtonen
Department of Military Technology, National Defence University, Helsinki, Finland, and

Kai Virtanen
Department of Mathematics and Systems Analysis, School of Science, Aalto University, Helsinki, Finland and Department of Military Technology, National Defence University, Helsinki, Finland

Abstract

Purpose – The purpose of this paper is to propose and test a multi-criteria decision analysis (MCDA) approach based on an additive value function (AVF) to select the most economically advantageous tender under European Union public procurement regulations.

Design/methodology/approach – A case study in which the AVF tender evaluation model is constructed by the procurement personnel and the results of the original, real-life public procurement evaluation model are compared to those discovered by the MCDA approach.

Findings – The AVF model captures the preferences of the procurement authority in a more reliable and transparent manner than commonly used evaluation models based on scoring formulas.

Practical implications – While commonly used in public procurement, relative scoring formulas can neither present the preferences of a procurement unit accurately nor do they enable bidders to draft bids according to these preferences. The proposed MCDA approach can achieve both.

Originality/value – The contribution of this paper is threefold. First, the successful construction of the AVF model with procurement personnel is introduced. Second, the model is used in an actual, real-life case. Third, a thoughtful comparison of features, structures and results of the AVF model and the evaluation model using scoring formulas is presented.

Keywords Procurement, Purchasing, Award criteria, Multi-criteria decision analysis

Paper type Research paper

1. Introduction

Public procurement in the internal European Union (EU) markets is regulated by public procurement directives 2014/24/EU, 2014/25/EU, 2014/23/EU and the defence procurement directive 2009/81/EU. These regulate how public procurement is carried out in the EU common market. The overall objective of the current public purchasing directives is to "obtain better value for public money, to deliver better outcomes for societal and other public policy objectives while increasing the efficiency of public spending" (European Commission, 2017).

There are three types of conflicting goals in every public procurement system: the goal to achieve efficiency or value-for-money; the aim to achieve other general political objectives

Journal of Public Procurement
Vol. 22 No. 2, 2022
pp. 164-179

1535-0118
DOI 10.1108/JOPP-06-2021-0040

such as regional balance and eco-friendliness; and the goal to maintain general trust in public procurement activities through transparency, equality and regulation (Schapper *et al.*, 2006). For example, Keulemans and Van de Walle (2017) note that many policy objectives contrast sharply with the principle of economy and choose the cheapest bidder.

As the value of public purchasing in EU is around 14% of gross domestic product (EU, 2019), the significant amounts at stake make public procurement a potential lever for achieving social, environmental and innovation-related objectives (Saussier and Tirole, 2015 p. 7). For example, the current Finnish Government not only intends to increase the efficiency of public procurement but also sees public procurement as a means to achieve social, climate and sustainability targets and even to promote domestically sourced food (Programme of Prime Minister Sanna Marin's Government, 2019). The transaction costs of public procurement have been estimated by Strand *et al.* (2013), whose overall average for Europe was at 1.4% of purchasing volume but with considerable variation between different EU member states. An average public procurement process uses 123 person-days of resources; in monetary terms, that equates to 28,000 EUR (Strand *et al.*, 2013). Of that, three-fourth are borne by businesses who draw up the bids.

The overriding goal of public procurement is to fulfil the needs of a procurement authority (PA) through a purchase from the private sector. The goal of each procurement procedure is the selection of the best tender, and therefore, the evaluation criteria must help clarify and objectivate what constitutes the "best tender" from the PA's standpoint (Mateus *et al.*, 2010). Public articles 26–31 in the 2014/24/EU procurement directive offer six different procurement procedures for the public PA. In selecting the tender, the PA first applies the mandatory (e.g. non-payment of taxes) exclusion grounds in Article 57, the optional exclusion grounds in Article 57 (e.g. previous failure to deliver) and the selection criteria in Article 58 (e.g. economic standing) to the tenderer. Then, the PA checks if the tender is compliant with all the requirements and conditions of the procurement documents (e.g. technical specifications). Assessing the best tender is based on the contract award criteria in Article 67. These award criteria may consist of either the price only or both the price and quality criteria. The multiple criteria option is termed the *most economically advantageous tender* (MEAT) in the directive. Furthermore, Article 67 requires that in addition to award criteria, the PA must also provide their relative weighting in the procurement documents. Regardless of which procurement procedure the PA chooses to use, both the award criteria and their relative weighting need to be determined at the beginning and are not subject to negotiation.

When considering the evaluation of alternatives in respect to multiple criteria, according to Keeney (2002), valuing the trade-offs between the criteria in the absence of the consequence ranges of the alternatives is a mistake. The requirement for setting weights for the award criteria in the absence of the range of their measurement levels is not a meaningful value trade-off, and thus, such weights and criteria do not necessarily reflect the utility of the PA. That is, it is not meaningful to say that the price of a car is four times more important than its top speed, unless a range of consequences is specified, for example, prices of 15,000 EUR and 20,000 EUR and top speeds of 150 km/h and 200 km/h. However, the procurement directive 2014/24/EU requires only criteria and their weightings. Moreover, in its decision on TNS Dimarso NV v Vlaams Gewest, the European Court of Justice has ruled that the PA is not required in advance to inform the tenderers of "the method of evaluation used by the contracting authority in order to specifically evaluate and rank the tenders", while "the method may not have the effect of altering the award criteria and their relative weighting".

Multicriteria contractor selection must be understood as a system where criteria, weights, scales and price formulas interact (Waara and Bröchner, 2006). To implement an entire evaluation model, in addition to multiple criteria and their weightings, also a procedure to evaluate for each tender and criterion the relative attractiveness of each performance level and then to combine the results for each criterion to an overall tender score is needed. A common way to assess the relative attractiveness of performance levels is through a scoring formula and using a weighted average to combine scores into an overall tender score.

Stilger *et al.* (2017) present 38 different scoring formulas found in the literature and other sources. They point out that the choice of formula critically affects which bid wins. Indeed, it should come as no surprise that, given the same input values, different formulas produce different results.

Scoring formulas can be divided (Faustino, 2017 p. 24) into relative formulas, where the bid score is measured for a criterion in relation to other tenders (e.g. as a percent of best tender) and independent formulas, where the bid score for a criterion is measured using an independent scale (e.g. the percentage of some pre-set value). There is not a full picture available of the use of different scoring formulas in Europe, as the procurement directive does not mandate their publication. Moreover, each EU country may have national legislation, legal practices and administrative rules about tender evaluation models. Faustino (2017) suggest that in the UK and France, the use of relative formulas for price criteria are a common practice. Högnäs and Kortelainen (2019) mention that in Finland, relative formulas are widespread in practice and are the default in electronic tendering systems, while in Sweden, relative formulas are avoided. Portugal has even prohibited relative formulas through legislation (Mateus *et al.*, 2010).

Relative formulas have two clear issues recognised by a number of authors. One is rank reversal or the ranking paradox, where the order of two bids may be changed by presenting a third bid (Stilger *et al.*, 2017; Waara and Bröchner, 2006; Bergman and Lundberg, 2013; Chen, 2008), thus violating the rule independence of irrelevant alternatives. Stilger *et al.* (2017) report a Dutch court case where it was found to be fully legal that a scoring formula could produce a reversal between two bids when a third bid was excluded as non-compliant. The other is that, because of scoring in relation to other bids, a tenderer cannot calculate the score when preparing the bid and, therefore, cannot optimise their own bid in relation to their production costs (Bergman and Lundberg, 2013; Dini *et al.*, 2006; Mateus *et al.*, 2010). Relative scoring methods replace the preferences of a buyer to a certain extent with a lottery because the lowest price is determined by the market and not by the buyer (Faustino, 2017).

In addition to scoring formulas, *multi-criteria decision analysis* (MCDA) has also been proposed for the bid selection problem, for example, by Mateus *et al.* (2010) and Sciancalepore *et al.* (2011). In general, decision analysis (Clemen, 1996; Keeney and Raiffa, 1993; Kirkwood, 1997) is a discipline which formalises the analysis of decisions. The primary aim of decision analysis is to aid a decision-maker – or a group of decision-makers – to think systematically about their values, objectives and preferences, as well as to structure decision problems. Decision analysis offers a versatile set of methods to analyse decision problems in a transparent manner and to support decision-making.

Mateus *et al.* (2010) introduce an MCDA bid evaluation model based on an additive value function (AVF) (Keeney and Raiffa, 1993) originating from the multi-attribute value theory. Although the construction and use of this model in the bid selection are discussed in an elegant way by Mateus *et al.* (2010), there is no application to an actual tendering process, as only an illustrative example is presented. It should also be noted that the quality-to-price method introduced by Bergman and Lundberg (2013) for the bid selection can be seen as a special case of the value function when the value gained from a criterion is assumed to depend linearly on the measurement level of that criterion. However, with the AVF, the form of this dependence is not restricted to a linear relationship. This enables a more flexible way to describe preferences of the PA.

Like the value function, *analytical hierarchy process* (AHP) (Saaty, 2000) is a widely used tool in the field of MCDA. AHP-based tender evaluation models are introduced, by Ishizaka *et al.* (2012), Marcarelli and Nappi (2019) and Marcarelli and Squillante (2019). For instance, Marcarelli and Nappi (2019) use an ex-post evaluation approach where AHP is used to rank the tenders using the MEAT criteria. However, there are several arguments favouring the use of the value function over AHP in the bid selection. For instance, in AHP, relative pairwise comparisons of alternative bids are conducted, and thus, AHP suffers similar issues as the relative scoring formulas discussed above.

Outranking methods (Ishizaka and Nemery, 2013) have also been suggested for evaluating bids (Segura and Maroto, 2017), but compared to the AVF, one of their challenges is the vague interpretation of criteria weights that should represent the relative importance of performance differences measured with each criterion. Moreover, the PA should determine indifference and preference thresholds on the measurement scales of the criteria. The thresholds imply whether two bids are equally good or whether one bid is preferred to another one with respect to each criterion. In general, the elicitation of these weights and thresholds from the PA can be difficult. Moreover, the quantification of the weights is based on relative pairwise comparisons, which pose the undesirable threat of rank reversal to the bid evaluation process.

In this article, the selection of the MEAT under EU public procurement regulations is tackled from the perspective of MCDA. In the tender preparation stage, there are no actual bids, so the evaluation model including criteria as well as their weights and scoring formula corresponding to partial value functions of the criteria in the AVF model should be able to reflect the preferences of the PA. That is, bids that are more preferable to the PA should also receive higher overall scores. We propose an MCDA approach based on an AVF for tender preparation. In this approach, the PA, given the MEAT criteria, first determines partial value functions representing the PA's preferences related to each criterion by using a bisection procedure (Keeney and Raiffa, 1993). In the second phase of the approach, the PA assesses the weights of the criteria by using the SWING procedure (Von Winterfeldt and Edwards, 1986). The novelty in comparison to the existing literature and especially to Mateus *et al.* (2010) is that this article presents an ex-post application of the approach with a real public procurement tendering case of the Finnish Defence Forces (FDF). In addition to testing the feasibility of the proposed approach, an application with a real ex-post case also enables a comparison to results obtained with the actual tender evaluation model and, thus, for conclusions to be drawn on the merits of both evaluation means, including the amount of elicitation effort and expertise required from the PA.

2. The multi-criteria decision analysis approach for tender evaluation

In the MCDA approach, an AVF is used to provide overall scores of tenders. Its elements are determined by the PA using decision analytic elicitation procedures. As discussed in the introduction section, such a value function is a widely used method for analysing multi-criteria decision-making problems, and Mateus *et al.* (2010) have also suggested its use in evaluating tenders and identifying the MEAT in public procurement processes.

The AVF is composed of partial scores of the tenders with respect to each award criterion and weights of the award criteria. Formally, it is of the form

$$V(x) = \sum_{i=1}^{n} w_i v_i(x_i) \tag{1}$$

where V(x) is the overall score of tender x, $w_i \geq 0$ is the weight, $v_i(x_i)$ is the partial value function of award criterion i and x_i is the measurement level of tender x on the measurement scale of criterion i. The value range of the partial value functions is [0, 1]. The worst measurement level of each criterion among the set of tenders to be compared is associated with score 0 and the best one with score 1 and other measurement levels between the worst and best levels are assigned to a score between 0 and 1. Therefore, the original measurement scales of the award criteria are converted into commensurate unit scales using these functions. The weights are normalised so that their sum is 1. Because of the scaling of the partial value functions and the normalisation of the weights, the overall scores of the alternative tenders vary within the range from 0 to 1. These scores provide the final rank of the tenders and a tender with the highest overall score wins the bidding competition. Note that the AVF reveals a natural interpretation of the weights of the award criteria associated with the ranges of the worst and best measurement levels; the weight of the given criterion is equal to the change of the overall score of a tender when the measurement level of this criterion moves from the worst level to the best level. Furthermore, the ratio of two weights reflects the trade-off between the partial scores of the corresponding criteria.

Regarding the theoretical rationale of the AVF, its functional form is motivated and argued by underlying preference assumptions or axioms. That is, when the PA accepts these axioms, there exists an AVF that captures the PA's preferences. From the practical point of view, the most relevant axiom deals with the preference independence of award criteria. Roughly speaking when the preference order for the measurement levels of one criterion does not depend on the level of another criterion, the criteria are said to be mutually preferentially independent (Keeney and Raiffa, 1993). Then, the AVF is a valid representation of the preferences of the PA. If the criteria are not preference independent in this sense, then more complex value functions, for example, a multiplicative function, should be used (Dyer and Sarin, 1979). A way to overcome the preference independence issue is to define a new preference independent criterion whose measurement scale is two-dimensional, representing the scales of two originally preferentially dependent criteria (Ewing and Parnell, 2006).

2.1 Elicitation of partial value functions

The decision analysis literature offers several procedures for eliciting partial value functions and weights of award criteria from the PA. The choice of an assessment procedure of the partial value functions depends on the type of measurement scales of the criteria (Eisenfuhr *et al.*, 2010). For example, if an award criterion has an interval or ratio scale, then the difference standard procedure can be used. In this procedure, a sequence of equally preferred differences on the measurement scale of a criterion should be assessed by the PA. Based on these assessments, a piecewise linear partial value function can be determined (von Winterfeldt and Edwards, 1986). In some cases, it might be appropriate to use a pre-specified functional form, for example, exponential, for the partial value function (Kirkwood, 1997; Mateus *et al.*, 2010). In the case of an ordinal measurement scale, the direct assessment of measurement levels can be applicable in providing partial score for each level (Clemen, 1996). Moreover, scores could also be determined based on pairwise comparisons between the measurement levels. Indeed, such comparisons are used to express preference information when applying AHP in the selection of the MEAT (Ishizaka *et al.*, 2012; Marcarelli and Nappi, 2019; Marcarelli and Squillante, 2019).

In the MCDA approach, partial value functions are elicited from the PA using a bisection procedure (Keeney and Raiffa, 1993). This is a suitable technique for the case study presented in this article, as all the award criteria are measured on ratio scales and its use is already illustrated in tender evaluation models (Mateus *et al.*, 2010). The popularity of this procedure in MCDA studies is mainly based on a pragmatic consideration that decision-makers typically prefer applying the bisection procedure rather than using, for example, the difference standard procedure (Schuwirth *et al.*, 2012; Ferretti, 2016; Zheng and Lienert, 2012; Zheng *et al.*, 2016). It has also been noted that the bisection procedure provides more reliable partial value functions compared to, for example, ones elicited via direct rating (Schuwirth *et al.*, 2012).

In the bisection procedure, the PA should choose a midpoint between two measurement levels of a single criterion so that the improvement of the score from the one measurement level to the midpoint is equal to the improvement of the score from the midpoint to the other measurement level. This gives two new intervals of measurement levels whose midpoints are identified in a similar way. The procedure is repeated until a set of (measurement level, score)-points is obtained which allows the definition of the partial value function with the desired accuracy.

2.2 Elicitation of criteria weights

There are several procedures for assessing weights of the AVF that represent preferences of the PA. Mateus *et al.* (2010) discuss the use of alternative elicitation procedures in the bid selection but only two of them, that is, the trade-off procedure (Keeney and Raiffa, 1993) and the SWING procedure (Von Winterfeldt and Edwards, 1986), take into account the ranges of measurement levels of the award criteria which is essential for the resulting weights to reflect the preferences of the PA. Direct weighting procedures, such as SMART (Edwards, 1977) or pairwise comparisons in AHP, ignore these measurement level ranges leading to weights with no meaningful interpretation. There are also other arguments for the superiority of the SWING and trade-off procedures over AHP pairwise comparisons in eliciting weights. The number of the PA's comparison tasks increases exponentially with the number of the criteria and the PA also must ensure that comparison results are consistent. On the other hand, experimental studies have shown that weights obtained using AHP typically deviate from SWING and trade-off weights (Pöyhönen and Hämäläinen, 2001). In addition to the ignorance of measurement level ranges in relative pairwise comparisons, vague interpretations of AHP weights also originate from the lack of well-founded underlying theory, that is, there is no preference axiomatic basis for AHP. Moreover, the mapping from the standard measurement scale used in AHP comparisons to the resulting weights is non-linear, which complicates how the levels of the scale and their verbal expressions can be understood by the PA. To overcome this challenge, alternative scales for pairwise comparisons have been suggested (Salo and Hämäläinen, 1997). Finally, AHP suffers a well-known rank reversal issue – the rank of tenders may change when a new tender is added to or an existing tender is excluded from the tender evaluation.

There are also weighting procedures that allow for the use of only ordinal or incomplete preference information to rank the importance of award criteria. In the former procedures, the PA should first rank the criteria. Then, weights are obtained based on this rank information (Bana e Costa *et al.*, 2002; Ahn and Park, 2008; Pictet and Bollinger, 2008). In the latter procedures, the preference information of the PA is expressed in the form of inequalities related to weights which define intervals for the weights (Weber, 1987; Salo and Hämäläinen, 1992; Harju *et al.*, 2019). Then, it is possible to determine the overall interval scores of the tenders, and these imply the ranking of the tenders.

The SWING procedure is selected for the case study because of its acknowledged means of providing weights in a consistent and valid way. The trade-off weighting procedure is often suggested as a preferable means for eliciting weights in the MCDA literature, and the use of this procedure is also described in tender evaluation models with a tokenistic example by Mateus *et al.* (2010). However, in trade-off weighting, the PA should construct and consider equally preferred hypothetical tender alternatives, which makes its use more time-consuming and cognitively demanding compared to the SWING procedure (Schuwirth *et al.*, 2012;

Fischer, 1995; Eisenfuhr *et al.*, 2010; Danielson and Ekenberg, 2019; Riabacke *et al.*, 2012). Moreover, before the use of SWING, only measurement level ranges of the award criteria are required, whereas shapes of partial value functions of the criteria are required and must be taken into account when applying the trade-off procedure (Schuwirth *et al.*, 2012; Eisenfuhr *et al.*, 2010; Ferretti, 2016). Therefore, the SWING procedure allows conducting weight elicitation before the assessment of the partial value functions. The SWING procedure has also proven to provide similar weights in test–retest experiments (Bottomley and Doyle, 2001), stable weights over time in a time-dependent decision environment (Lienert *et al.*, 2016) and has been shown to avoid some cognitive biases (Montibeller and von Winterfeldt, 2015), even though all weighting procedures have the potential for such biases (Hämäläinen and Alaja, 2008). Moreover, although SWING is less theoretically defensible than trade-off weighting, the convergence validity of these methods has been confirmed by demonstrating that the procedures provide similar weights in experimental settings (Pöyhönen and Hämäläinen, 2001).

In the SWING procedure, the PA should first consider a hypothetical tender in which all the criteria are their worst measurement level. The PA should choose the criterion that the PA would first like to change to its most preferred measurement level. This criterion is given 100 points. Then, the PA identifies the next criterion that the PA would like to change to its most preferred measurement level. This criterion is given points that reflect this improvement relative to the first one. The procedure continues in the same way until points of all the award criteria are determined. Finally, the points are normalised, that is, divided by their sum, which provides the weights of the AVF.

3. The mini unmanned aerial system procurement case

The MCDA approach was applied in an ex-post workshop to the procurement of a mini unmanned aerial system (MUAS). The actual MUAS procurement had already been completed by the FDF at a cost of €30m. The ex-post design makes it possible to compare the results of the MCDA approach with the actual evaluation model included in the request for quotations (RFQs). Because the procurement has been completed, the evaluators at the ex-post workshop were subject matter experts concerning this procurement and knowledgeable of the procurement needs. As the MUAS procurement was conducted by the FDF, the actual names of the MUAS systems are referred to by code names.

3.1 Mini unmanned aerial system tenders and request for quotations evaluation model

The MUAS procurement was carried out according to the defence procurement directive 2009/81/EC. The contract award criteria were price, quality, field test performance and life-cycle costs (LCCs). The quality criterion was measured as a sum of points given based on a number of optional MUAS features and performance criteria. The weights for the award criteria were specified in the RFQs. As required by Finnish legal practice, scoring formulas for the criteria were specified in the RFQs. The scoring formula of a criterion converts the criterion's measurement levels into partial scores of tenders in respect to that criterion. The RFQs evaluation model, including the scoring for each criterion and criteria weights, was

$$P_{tot} = \frac{P_l}{P_i} * 60 + \frac{Q_i}{Q_h} * 40 + \frac{F_i}{F_{max}} * 30 + \frac{LCC_l}{LCC_i} * 10 \quad (2)$$

where:

P_{tot} = overall score of a tender i;
P_l = lowest quoted price;
P_i = price of tender i;

Q_i = quality points of tender i;
Q_h = highest quality points;
F_i = field test points of tender i;
F_{max} = maximum field test points (30 p);
LCC_l = lowest quoted LCCs; and
LCC_i = LCC of tender i.

The scoring formulas of the award criteria were relative, except for the absolute measurement scale from 0 to 30 performance points for the field test criterion.

FDF received five tenders, one of which did not fulfil the selection criteria, so four tenderers were evaluated, including the extensive field testing done by the FDF. The four tenders and their levels on the measurement scales of the criteria are shown in Table 1.

3.2 Applying the multi-criteria decision analysis approach to tender evaluation at the ex-post workshop

A half-day workshop was arranged 24th March 2020 to apply the MCDA approach to tender evaluation. Because of the travel restrictions because of the coronavirus pandemic, the workshop was organised both face-to-face and virtually via a video link. The workshop participants included five persons, a technical expert and a commercial specialist who had drafted the original RFQ as well as another commercial specialist, a participant with legal expertise and a high-ranking manager.

After explaining the purpose of the workshop, the MCDA approach was introduced to all participants by the workshop facilitators (i.e. the authors of this paper). Explaining the bisection and SWING procedures was important to ensure sufficient knowledge for their use in the workshop. Regarding the use of the bisection and SWING procedures, they were applied in the workshop following the steps discussed in their general descriptions on Sections 2.1 and 2.2. The authors facilitated the elicitation of each partial value function included in the AVF model by explaining the meaning of choices in the bisection procedure, as the participants did not have prior experience of partial value functions or the bisection procedure. The set of (measurement level, score)-points obtained during the procedure and the resulting functions for all award criteria are presented in Figure 1. Eliciting the partial value functions lasted about 20 min for each of the four criteria and terminated once a consensus through group discussion was reached. After that, the criteria weights were elicited with the SWING procedure, and this terminated once a consensus was reached through group discussion in about 15 min.

For the price criterion, the participants arrived at a concave value function (Figure 1a). This reflected a strong preference to not exceed the budget constraint for the procurement. From the whole-of-the-government perspective, the value function should essentially be linear, as the cost of the MUAS procurement from that perspective is small. However, as Bergman and Lundberg (2013) point out, the PA may have weak incentives not to use all of

Table 1. Mini unmanned aerial system tenders, their measurement levels and the best and worst measurement levels of the award criteria

Criterion	Green	Blue	White	Black	Best	Worst
Price (M€)	10.5	29.6	25.1	12.6	10.5	29.6
Quality (points)	1 074	1 053	926	1 028	1 074	926
Field test (points)	20	19	19,5	16	20	16
LCC (k€/yr.)	486	559	726	500	486	726

the allocated funds and often strong incentives not to exceed the budget allocation causing the value function to be curved. On the other hand, for the quality criterion, the elicitation procedure revealed a convex value function that reflects the strong preference to achieve the maximum performance. The workshop participants indicated that achieving the best performance was a top priority. They arrived at a linear value function for the field test criterion (Figure 1c). The main purpose of the field test was to exclude MUAS tenderers that did not fulfil the minimum requirements. Therefore, the extra performance points were linear. There was some discussion of what shape the value function should be for the LCCs, but in the end, consensus for a linear value function was reached (Figure 1d). The verification issue of the LCCs was also brought up in the discussion, as contractual mechanisms for making an LCC complaint were weaker than that for other criteria. However, the participants decided that this should not be reflected in the value function.

When one inserts the best measurement levels (Table 1) of the criteria measured with relative scoring formulas, that is, the price, quality and LCC criteria, into the formulas included in equation (2), the corresponding partial scores can be plotted as a function of the measurement scale (Figures 1a, 1b and 1d). While these best levels are not known when the bidders draft their bids, they are available ex-post when all the bids have arrived. As the field test criterion has an independent scoring formula, the corresponding partial scores are readily available (Figure 1c).

When comparing the partial value functions and the RFQs scoring formulas presented in Figure 1, only with regard to the field test criterion, the workshop arrived at a partial value function with a similar linear form as in the RFQs evaluation model. For the price criterion, the value function was concave, while for the RFQs, it was a convex formula. For the quality criterion, the value function was strongly convex, while the RFQs formula was linear. The RFQs LCCs scoring formula was slightly convex, while the linear partial value function for the LCCs was elicited at the workshop.

When the SWING procedure for eliciting the criteria weights of the AVF evaluation model was executed at the workshop, first the importance points (see the points column in Table 2) regarding the ranges of the criteria measurement levels shown in Table 1 were

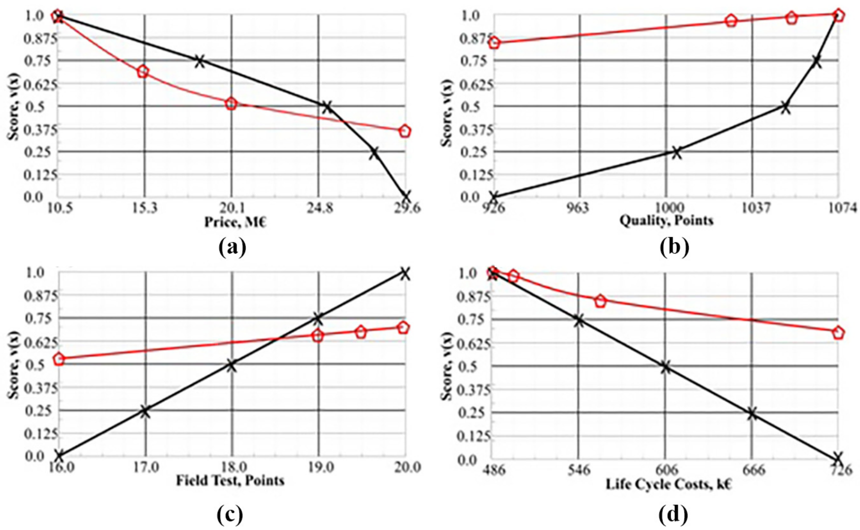

Figure 1.
Score of (a) price, (b) quality, (c) field test and (d) LCC criteria provided by the partial value function (black curve) and by the RFQ scoring formula (red curve). Black crosses refer to (measurement level, score)-points identified in the elicitation procedure. Red diamonds refer to the tenders

assessed. Then, the normalisation of these point resulted in the criteria weights presented in the weight column of Table 2. The corresponding RFQs weights are given in the normalised weight column of Table 3. Here, these weights were obtained by normalising the original weights of the RFQs evaluation model [equation (2)] so that their sum was 1. When the SWING weights are compared to the RFQs weights, a number of differences can be seen. The SWING procedure gives only around half the weight for the price criterion (0.20) than the RFQs model (0.43), while the SWING weight for quality (0.40) has increased the most compared to its RFQs weight (0.29). The weight given to the field test criterion is almost unchanged, while the LCCs weight has increased.

Because the range of the partial score provided by the partial value function of each criterion is one, the SWING weights (Table 2) directly describe the contribution of each criterion to the overall score differences between the tenders under consideration when using the AVF evaluation model. To analyse the contributions of the criteria also in the case of the RFQs evaluation model, the product of each criterion's normalised weight and score range was calculated, and the resulting products were normalised to sum them to 1. These quantities measure the effect of each criterion on the overall score differences between the tenders in the RFQs model. They are shown in the normalised product column of Table 3. When comparing the overall contributions of the criteria to the tender evaluation in the AVF model and in the RFQs model, the contribution of the price criterion in the RFQs is overwhelming, about 75% of total, while it is only around 20% with the AVF model. The rank order of the effects of the three non-price criteria remains the same with both evaluation models, although their magnitudes are clearly different.

The price criterion dominates the tender evaluation based on the RFQs model and does that even more so than could be concluded by analysing only the value of the RFQs weight for price. This finding clearly demonstrates that – as argued in the introduction section – not taking into account the ranges of the measurement levels of award criteria when assessing their weights and, on the other hand, using relative scoring formulas can lead to a biased evaluation result, including meaningless overall scores for tenders. By assuming that the AVF model constructed in the ex-post workshop reflects the preferences of the PA and these

SWING	Points	Weight
Price	50	0.20
Quality	100	0.40
Field test	60	0.24
LCC	40	0.16

Table 2. SWING points and weights of the award criteria

RFQ	Weight	Normalised weight	Highest score	Lowest score	Score range	Normalised weight × score range	Normalised product
Price	60	0.43	1	0.35	0.65	0.28	0.754
Quality	40	0.29	1	0.86	0.14	0.04	0.11
Field test	30	0.21	0.67	0.53	0.13	0.03	0.074
LCC	10	0.07	1	0.67	0.33	0.02	0.062

Table 3. Request for quotations weights of the award criteria and normalised product quantities describing the contribution of each criterion to the overall evaluation of the tenders

have remained constant since the actual MUAS procurement, the RFQs model poorly captures these preferences.

The overall scores of the tenders provided by the AVF model and the RFQs model are presented in Figures 2 and 3, respectively. The contribution of each criterion to these scores is also illustrated. Note that the RFQs scores were now calculated by using the normalised RFQs weights presented in Table 3. This way it is possible to compare their magnitude to the AVF scores determined with the normalised SWING weights presented in Table 2.

In the actual MUAS system procurement procedure, the FDF awarded the contract to the tenderer "Green" that received the highest overall RFQs score (0.93). This tender had also the highest overall AVF score (1.00). In fact, "Green" is the best alternative in respect to each single award criterion. Therefore, it would be awarded the contract regardless of which evaluation model is used. Nevertheless, the superiority of "Green" over the other tenders was

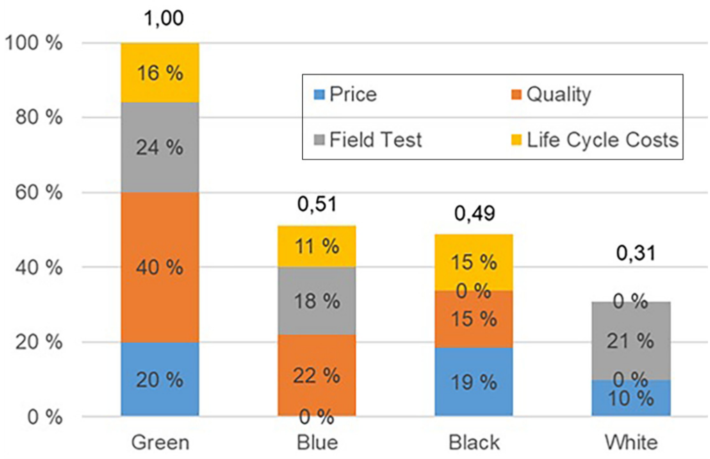

Figure 2.
Overall scores of the tenders provided by the AVF evaluation model. The score bars are divided according to the contribution of each criterion

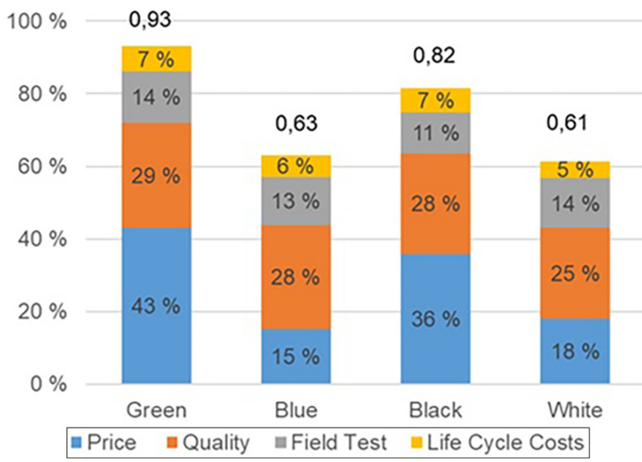

Figure 3.
Overall scores of the tenders provided by the RFQ evaluation model. The score bars are divided according to the contribution of each criterion

revealed more clearly with the AVF model than the RFQs model, as the AVF score of the second-best tender, that is, "Blue" was 0.51 and the RFQs score was 0.82 assigned to the tender "Black". That is, the evaluation models led to different ranks of these two tenders. Both models gave the worst rank to the tender "White" (AVF score 0.31 and RQFs score 0.61), but here also, the overall RFQs score was closer to the scores of the better tenders compared to the differences between the AVF scores. Overall, the sensitivity of the AVF scores seems to be more accurate than the RQFs scores, as the score range is noticeably wider with the AVF model (0.31–1.00) than with the RFQs model (0.61–0.93).

4. Discussion

The MCDA approach with an AVF was tested for the selection of the MEAT for the MUAS case. The workshop participants were able to elicit partial value functions using the bisection procedure and criteria weights with the SWING procedure. In both procedures, the participants' answers to elicitation questions were founded on ranges of criteria measurement levels. Compared to the RFQ approach in which criteria weights are decided and then a scoring rule for each criterion is selected, the MCDA approach brings more clarity and openness to the value decisions. We believe that both the elicited partial value functions and SWING weights more properly represent the preferences of the PA.

The partial value functions elicited in the ex-post workshop were quite different compared to scoring formulas used in the RFQs evaluation model (Figure 1). It is hard to see that the relative scoring formulas for price and LCCs in the RFQs would represent what actually gives value to the PA. RFQs scoring means that there are diminishing marginal penalties the higher the bid price is in relation to the lowest price. The mere use of scoring formulas does not seem to encourage thinking about what adds value to the procurement at hand. It is almost as if it brings about a shortcut to arriving at a decision. The partial value functions obtained with the bisection procedure, on the other hand, are more believable. The use of this procedure, however, should be trained thoroughly, especially if there are no facilitators in an elicitation session.

The weight of the quality criterion increased while the weight of the price criterion decreased when applying the SWING procedure compared to the original RFQ weights. Thus, this procedure more clearly reveals different contributions of quality and price than the RFQs in the overall evaluation results. Determination of the weights with the SWING procedure was based on a firm theoretical footing, while it is hard to say what the basis was for the RFQs weights, other than guesswork. Provided that the preferences of the PA were unchanged since the actual MUAS procurement, the SWING weights more correctly reflect the preferences of the PA than the RFQs weights.

Both AVF and RFQs evaluation models ranked the same choice first, which was natural, as the selected tender was the best one in respect to all the award criteria, that is, it dominated all other alternatives. It is important to note, however, that the evaluation models could have led to different first choices if the tenders had been more even. In fact, the rank was changed between Black and Blue, that is, the second- and third-best tenders.

If the original RFQs would have included partial value functions in Figure 1 and SWING weights in Table 2 instead of the evaluation model [equation (2)], the bidders could have calculated trade-offs in design and production and maximised their overall evaluation score. This calculation would have also maximised the value for the PA. In fact, it would have been in the interest of each bidder to submit a tender that maximises the PA value, given their profit requirement.

In the MUAS case, the ex-post workshop used the actual tenders to elicit partial value functions and criteria weights. This is not possible in the ex-ante drafting of an RFQ, as the tenders will come after the RFQs. In defence procurement, a request for information (RFI), including price and product features, before drafting the RFQ is common practice. The directive 2014/24/EU Article 40 on preliminary market consultations allows an RFI and other such procedures. Instead of actual tenders, the PA may use information received in preliminary market consultations to make up plausible tenders that cover the expected range of each criterion's measurement scale and apply the MCDA approach as with actual tenders. When partial value functions are elicited with bisections and weights with the SWING procedure, only the criteria and their expected range of measurement scales are needed.

Regardless of how well preliminary market consultations are carried out, there is always a possibility that the measurement levels of some criteria in some bids are over or below the expected range. If, for example, the price is lower than the expected price, then awarding no extra scores is one possibility, but it is not an optimal one from the PA perspective. Another possibility would be to extrapolate partial value functions outside the expected ranges, but it is unclear how courts would react to this. Faustino (2017, p. 154) reports a decision by a French court where negative scores were not allowed because it was thought that the impact of those negative scores on the overall score of each tender might distort the relative weighting of the criteria as disclosed in advance. However, it should be noted that there is no decision theoretical restriction that a partial value function could give scores greater than 1 or less than 0 when a measurement level is outside the expected range.

The weight assigned to the quality criterion in the ex-post workshop was much higher than that of the RFQs. The RFQs included over 100 individual features and performance requirements that each gave some amount of points. The quality points used in the RFQs evaluation model [equation (2)] was a sum of such individual items fulfilled by a tender. Even the worst tender fulfilled around six-seventh of the those fulfilled by the best tender. If this would not have been found in the preliminary market consultations, then the PA may have had to draft a partial value function covering the entire range of the quality scale. In that case, it may not be assumed that the shape of the value function would have been similar to the shape identified at the workshop. For example, would the form of red line in Figure 1b have remained the same in the interval [926, 1074], if the range [0, 1074] instead of [926, 1074] had been used as a basis in the workshop? The obtained convex form reflected the strong preference for procuring the best possible equipment, but it also highlights the need for educating the procurement personnel in thinking about and eliciting what it is that provides value to the PA for the procurement. In the ex-post workshop, the partial value function of the price criterion strongly implied the importance of not exceeding a budget constraint and small incentives for not using all the money. From the overall governmental perspective, the partial value function of the price should be linear, as the saved money could be used elsewhere. In fact, as the PA does not have a budget constraint for LCCs, its value function was linear.

5. Conclusions

There is a large body of literature on decision analysis approaches for making value decisions. However, as pointed out by Bergman and Lundberg (2013), the current public procurement practice and practitioners are not informed by this literature. The issues with widely used tender evaluation models based on scoring formulas are that:

- Their capability to capture preferences of the PA is questionable as the preference representation depends on the bids in the evaluation, not on the view of the PA.
- Weights of award criteria are assessed without taking into account ranges of measurement levels of the award criteria, leading to vague interpretation and meaning of the weights.

- Relative scoring formulas do not enable bidders to make informed judgements regarding their offer because the scoring depends on other bids that are unknown at the time.

Therefore, the nature of these evaluation models can be viewed as a black box, and the evaluation process may end up with a contract that is not made based on well-argued rationales reflecting the original aims of the PA. To overcome the issues discussed, we propose an MCDA approach with an AVF that could be applied in public procurement tendering. Although the application of AVF-based models to support the tender evaluation has been earlier suggested in the literature (Mateus *et al.*, 2010), our article is the first in which the successful construction of the AVF model with subject matter experts (i.e. the procurement personnel) is introduced and in which such a model is used in an actual, real-life case. In addition, a thoughtful analysis and comparison of are conducted of the features, structures and results of the AVF model and the evaluation model using commonly used scoring formulas.

The managerial practice of public procurement could benefit from clear thinking about what really brings value to the PA. The application of the MCDA approach took only 2 h from four participants, that is, one person-day (excluding the researchers) in an ex-post workshop. From a managerial viewpoint, the effort of applying the proposed approach is only a fraction of the effort required for a public procurement.

References

Ahn, B. and Park, K. (2008), "Comparing methods for multi attribute decision making with ordinal weights", *Computers and Operations Research*, Vol. 35 No. 5, pp. 1660-1670.

Bana e Costa, C., Correa, E., Corted, J. and Vansnickd, J.C. (2002), "Facilitating bid evaluation in public call for tenders: a socio-technical approach", *Omega*, Vol. 30 No. 3, pp. 227-242.

Bergman, M.A. and Lundberg, S. (2013), "Tender evaluation and supplier selection methods in public procurement", *Journal of Purchasing and Supply Management*, Vol. 19 No. 2, pp. 73-83.

Bottomley, P. and Doyle, J. (2001), "A comparison of three weight elicitation methods: good, better and best", *Omega*, Vol. 29 No. 6, pp. 553-560.

Chen, T.H. (2008), "An economic approach to public procurement", *Journal of Public Procurement*, Vol. 8 No. 3, pp. 407-430.

Clemen, R.T. (1996), *Making Hard Decisions*, Wadsworth Publishing Company, Belmont, CA.

Danielson, M. and Ekenberg, L. (2019), "An improvement to swing techniques for elicitation in MCDM methods", *Knowledge-Based Systems*, Vol. 168, pp. 70-79.

Dini, F., Pacini, R. and Valletti, T. (2006), "Scoring rules", in Dimitri, N., Piga, G. and Spagnolo, G. (Eds), *Handbook of Procurement*, Cambridge University Press, New York, NY, pp. 293-319.

Dyer, J.S. and Sarin, R.K. (1979), "Measurable multi attribute value functions", *Operations Research*, Vol. 27 No. 4, pp. 810-822.

Edwards, W. (1977), "How to use multi attribute utility measurement for social decision making", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 7 No. 5, pp. 326-340.

Eisenfuhr, F., Weber, M. and Langer, T. (2010), *Rational Decision Making*, Springer Verlag, Berlin.

European Commission (2017), "Making public procurement work in and for Europe", available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52017DC0572&from=EN (accessed 21 December 2019).

European Union (2019), "Public procurement", available at: https://ec.europa.eu/growth/single-market/public-procurement_en (accessed 21 December 2019).

Ewing, P.L. and Parnell, G.S. (2006), "Use of decision analysis in the army base realignment and closure (BRAC) 2005 military value analysis", *Decision Analysis*, Vol. 3 No. 1, pp. 33-49.

Faustino, P.B. (2017), "Public procurement award criteria in the EU member states: a comparative study of the discretionary powers of contracting authorities", PhD thesis, University of Nottingham.

Ferretti, V. (2016), "From stakeholders analysis to cognitive mapping and multi-attribute value theory: an integrated approach for policy support", *European Journal of Operational Research*, Vol. 253 No. 2, pp. 524-541.

Fischer, G. (1995), "Range sensitivity of attribute weights in multi attribute value models", *Organizational Behavior and Human Decision Processes*, Vol. 62 No. 3, pp. 252-266.

Hämäläinen, R. and Alaja, S. (2008), "The threat of weighting biases in environmental decision analysis", *Ecological Economics*, Vol. 68 Nos 1/2, pp. 556-569.

Harju, M., Liesiö, J. and Virtanen, K. (2019), "Spatial multi-attribute decision analysis: axiomatic foundations and incomplete preference information", *European Journal of Operational Research*, Vol. 275 No. 1, pp. 167-181.

Högnäs, E.-R. and Kortelainen, A. (2019), "Hinnan suhteellinen vertailu julkisessa hankinnassa (the relative comparison of price in public procurement)", available at: www.edilex.fi/artikkelit/19627 (accessed 13 March 2021).

Ishizaka, A. and Nemery, P. (2013), *Decision Analysis: Methods and Software*, John Wiley and Sons, Chichester.

Ishizaka, A., Pearman, C. and Nemery, P. (2012), "AHPSort: an AHP-based method for sorting problems", *International Journal of Production Research*, Vol. 50 No. 17, pp. 4767-4784.

Keeney, R.L. (2002), "Common mistakes in making value trade-offs", *Operations Research*, Vol. 50 No. 6, pp. 935-945.

Keeney, R.L. and Raiffa, H. (1993), *Decision with Multiple Objectives: Preferences and Value Tradeoffs*, Cambridge University Press, New York, NY.

Keulemans, S. and Van de Walle, S. (2017), "Cost-effectiveness, domestic favouritism and sustainability in public procurement", *International Journal of Public Sector Management*, Vol. 30 No. 4, pp. 328-341.

Kirkwood, C.W. (1997), *Strategic Decision Making*, Wadsworth Publishing Company, Belmont, CA.

Lienert, J., Duygan, M. and Zheng, J. (2016), "Preference stability over time with multiple elicitation methods to support wastewater infrastructure decision-making", *European Journal of Operational Research*, Vol. 253 No. 3, pp. 746-760.

Marcarelli, G. and Nappi, A. (2019), "Multi criteria approach to select the most economically advantageous tender: the application of AHP in Italian public procurement", *Journal of Public Procurement*, Vol. 19 No. 3, pp. 201-223.

Marcarelli, G. and Squillante, M. (2019), "A group-AHP based approach for selecting the best public tender", *Soft Computing*, Vol. 24 No. 18, pp. 13717-13724.

Mateus, R., Ferreira, J.A. and Carreira, J. (2010), "Full disclosure of tender evaluation models: background and application in Portuguese public procurement", *Journal of Purchasing and Supply Management*, Vol. 16 No. 3, pp. 206-215.

Montibeller, G. and von Winterfeldt, D. (2015), "Cognitive and motivational biases in decision and risk analysis", *Risk Analysis*, Vol. 35 No. 7, pp. 1230-1251.

Pictet, J. and 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*, Vol. 185 No. 3, pp. 1300-1307.

Pöyhönen, M. and Hämäläinen, R.P. (2001), "On the convergence of multi attribute weighting methods", *European Journal of Operational Research*, Vol. 129 No. 3, pp. 569-585.

Programme of Prime Minister Sanna Marin's Government (2019), *Inclusive and Competent Finland – A Socially, Economically and Ecologically Sustainable Society*, Publications of the Finnish Government, Helsinki, 10 December 2019.

Riabacke, M., Danielson, M. and Ekenberg, L. (2012), "State-of-the-art prescriptive criteria weight elicitation", *Advances in Decision Sciences*, Vol. 2012, pp. 1-24.

Saaty, T. (2000), *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*, RWS Publications, Pittsburgh, PA.

Salo, A. and Hämäläinen, R.P. (1992), "Preference assessment by imprecise ratio statements", *Operations Research*, Vol. 40 No. 6, pp. 1053-1061.

Salo, A. and Hämäläinen, R.P. (1997), "On the measurement of preferences in the analytic hierarchy process", *Journal of Multi-Criteria Decision Analysis*, Vol. 6 No. 6, pp. 309-319.

Saussier, S. and Tirole, J. (2015), "Strengthening the efficiency of public procurement", *Notes du Conseil D'analyse* Économique, Vol. 22 No. 3, pp. 1-12.

Schapper, P., Malta, J. and Gilbert, D. (2006), "An analytical framework for the management and reform of public procurement", *Journal of Public Procurement*, Vol. 6 Nos 1/2, pp. 1-26.

Schuwirth, N., Reichert, P. and Lienert, J. (2012), "Methodological aspects of multi-criteria decision analysis for policy support: a case study on pharmaceutical removal from hospital wastewater", *European Journal of Operational Research*, Vol. 220 No. 2, pp. 472-483.

Sciancalepore, F., Falagario, M., Cosentino, N. and Pietroforte, R. (2011), "Multi-criteria bid evaluation of public projects", Paper Presented at the Management and Innovation for a Sustainable Built Environment, Delft University of Technology, Amsterdam, 20–23 June.

Segura, M. and Maroto, C. (2017), "A multiple criteria supplier segmentation using outranking and value function methods", *Expert Systems with Applications*, Vol. 69 No. 1, pp. 87-100.

Stilger, P.S., Siderius, J. and Raaij, E.M.V. (2017), "A comparative study of formulas for choosing the economically most advantageous tender", *Journal of Public Procurement*, Vol. 17 No. 1, pp. 89-125.

Strand, I., Ramanda, P. and Canton, E. (2013), "Public procurement in Europe – cost and effectiveness, prepared for the European commission by PwC, London economics and ecorys", available at: https://op.europa.eu/en/publication-detail/-/publication/0cfa3445-7724-4af5-8c2b-d657cd690c03 (accessed 10January 2021).

Von Winterfeldt, D. and Edwards, W. (1986), *Decision Analysis and Behavioural Research*, Cambridge University Press, Cambridge, MA.

Waara, F. and Bröchner, J. (2006), "Price and non-price criteria for contractor selection", *Journal of Construction Engineering and Management*, Vol. 132 No. 8, pp. 797-804.

Weber, M. (1987), "Decision making with incomplete information", *European Journal of Operational Research*, Vol. 28 No. 1, pp. 44-57.

Zheng, J. and Lienert, J. (2012), "Stakeholder interviews with two MAVT preference elicitation philosophies in a Swiss water infrastructure decision: aggregation using SWING-weighting and disaggregation using UTA", *European Journal of Operational Research*, Vol. 267 No. 1, pp. 273-287.

Zheng, J., Egger, C. and Lienert, J. (2016), "A scenario-based MCDA framework for wastewater infrastructure planning under uncertainty", *Journal of Environmental Management*, Vol. 183 No. 3, pp. 895-908.

Corresponding author
Juha-Matti Lehtonen can be contacted at: juha-matti.lehtonen@mil.fi