Common Policy Appraisal Format

Structured Decision Evaluation of Policy Options using the Simulation Results

Project acronym:	**SENSE4US**
Project full title:	**Data Insights for Policy Makers and Citizens**
Grant agreement no.:	**611242**
Responsible:	**Stockholm University - eGovLab**
Contributors:	**Aron Larsson, Osama Ibrahim**
Document Reference:	**D6.3**
Dissemination Level:	**PU**
Version:	**Final**
Date:	**16/10/2015**

Project co-funded by the European Union under the Seventh Framework Programme

History

Version	*Date*	*Modification reason*	*Modified by*
0.1	2015-09-01	First draft	Aron Larsson, Osama Ibrahim
0.2	2015-09-30	Second draft, for internal review	Aron Larsson, Osama Ibrahim
1.0	2015-10-16	Final	Aron Larsson, Osama Ibrahim

Table of Contents

List of Figures

List of tables

List of Abbreviations

<Abbreviation>	**<Explanation>**
DM	Decision Maker
DSS	Decision Support System
GUI	Graphical User Interface
EC	European Commission
IA	Impact Assessment
ICT	Information and Communication Technology
MADM	Multi Attribute decision-making
MCDA	Multi-Criteria Decision Analysis
Sense4us	Data insights for policy makers and citizens (this project)
ST	Stakeholder
WP	Work Package

Executive Summary

So far, we presented in (D6.1) our decision support framework for policy formulation and researched the feasibility and information requirements of the proposed modelling and simulation method for policy analysis. (D6.2) provided a prototype for the policy modelling and simulation tool that mines related data from several online and other related work-package sources. This deliverable (D6.3) defines a common policy appraisal format for use in decision modelling. In (D6.4) a preference elicitation method and a process for structured negotiation will be designed based upon the common appraisal format and presented in a prototype for the policy decision evaluation tool that elicits weights, priorities and preferences from policy makers and stakeholders Finally, (D6.5) will provide a prototype for a software that processes the simulation outputs enabling for policy makers and stakeholders to express preferential statements with respect to simulated consequences. It includes user interface design and visualization of the impacts of decisions relative to each of the target groups based on computational decision analysis.

In this deliverable, we deal with the design of a common format for evaluation of the competing policy options identified from the policy simulation results. The objective of developing criteria models and data formats for policy appraisal is to assess the relative performance of the alternative policy options against multiple and perhaps conflicting objectives or criteria. Multi-criteria decision analysis (MCDA) is used for defining the main evaluation criteria and the underlying measurable attributes.

While the scenario-based dynamic simulation of policy consequences allows decision makers to identify the feasible policy options and verifies their economic, social and environmental impacts over time, it does not provide the explicit evaluation of policy options. MCDA can support an in-depth performance evaluation of policy options and consequently it can support the design of better options.

An explicit evaluation process of policy options involves judging the effectiveness, efficiency, relevance, coherence and added value of the alternative policy options. This deliverable provides the process of developing a criteria model for evaluation of policy options in relation to the economic, social, environmental and other policy impacts. It shows how each criterion can be operationalised using measurable attributes that can be derived from the simulation results and the work done in WP5 (notably the evidence extraction). A clear definition is provided for each evaluation criterion, along with the process and degree of analysis conducted and how to construct a data format using the simulated policy impacts to judge the performance of the different policy instruments.

Further, by identifying the stakeholders (interest groups and target groups), we can elicit their evaluations and preferences based on the common policy appraisal format. Therefore, we can gain insights into the level of disagreement on policy proposals as well as how a policy could be efficiently refined for the mutual benefits of two or more stakeholder groups or arriving at more than one competing policy options for further decision analysis.

Introduction

Evaluation plays an important role in organisational learning, identifying, sharing and assessing practices. Thorough evaluation can identify unintended and unexpected consequences, which also need to be taken into account. Commitment to an evidence-based policymaking approach, on the EU, national or local levels, requires that evaluations should be based on sufficient data and opinions of actors and stakeholders. Policy evaluations provide an opportunity to receive feedback and requests for change from those who are directly or indirectly affected by policy decisions.

Traditionally, evaluations have been used primarily in the assessment of policies and programs that are already in place. This ex post or retrospective application is the one most frequently associated with evaluation. Less frequently, prospective (forward-looking) evaluation methodology has been used to assess ex ante the potential success of policies that are under consideration. The conventional approaches to prospective evaluations have ranged widely from relatively freewheeling “demonstrations” to highly controlled field experiments. These approaches usually face the problem of providing useful information on the potential for success of policies and programs to a decision-making process that may take no longer than a year or two from proposal to definitive action. If evaluations are to contribute to decisions about proposed new programs, the contribution should be accomplished through procedures that are relatively inexpensive, speak to each of the variety of proposals under consideration, and provide timely results.

The U.S. General Accounting Office (1990), introduced the Prospective Evaluation Synthesis (PES). PES is a prospective analysis anchored in evaluation concepts that takes into account ways in which the past is and is not likely to be similar to plausible future conditions. It involves conceptual, operational and empirical analyses, taken in the context of the future. The conceptual analyses results help focus the operational analyses. The operational analyses further scope the search for empirical findings. The empirical analyses can open both new conceptual and operational possibilities.

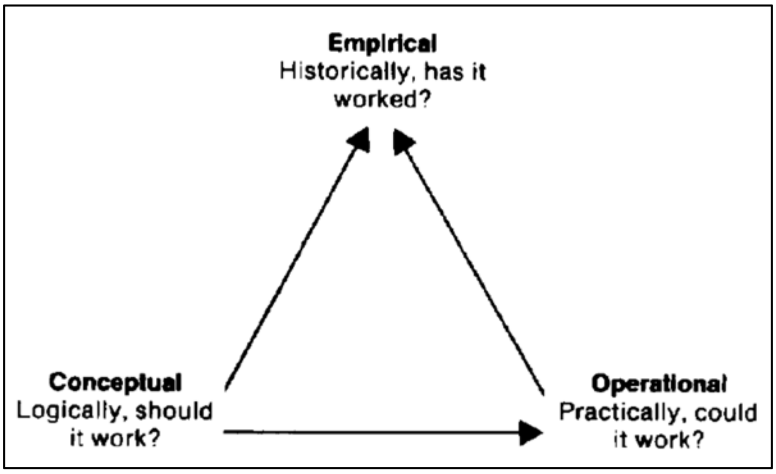

Figure 1 : The triad of analysis, PES

The work done in WP6 aims to support prospective and prescriptive policy analysis that shows the necessary actions to achieve targeted outcomes and the interrelated effects of a decision (Turnpenny et al. 2009). Similar to PES, the Sense4us policy modelling, simulation and decision analysis approach involves conceptual, operational and empirical analyses taken in the context of the future.

The conceptual analysis is represented in building a mental model of the policy or governance system that reflects what is known (or believed) about the target system. The modelling of public policy problems deals with complex systems usually characterised by insufficient knowledge, unresolvable uncertainties or absence of baselines (in some cases). Thus, the modeller has to make assumptions and estimations using expert's judgements or information from validated sources about details and mechanisms (e.g., Intensities and time lags of the causal links).

The empirical analysis involves exercising the model through scenario-based simulation to explore the implications of alternative actions and of varying assumptions and hypotheses, and it is suggested that the user collects as much data as possible to limit the number of computational experiments required to answer a policy question. The individual simulation runs are not being treated as providing predictions or explicit answers to the policy makers' questions. Instead, new information that was implicit in their prior knowledge that defined what is plausible, is generated to support an informed policy decision. Exploratory research strategies can provide a basis for decision making and a guide for information searching, through:

(i) identifying special/extremal cases;
(ii) suggesting plausible explanations for puzzling facts (hypotheses generation);
(iii) developing worst cases scenarios for risk aversion situations.

The model validation, which is to be done by the user, is then reduced to validation of sub-models, determining model parameters from validated sources and ensuring the plausibility of outcomes. In addition, the model quality control is limited to the plausibility and the degree of completeness of the model (inclusion of all factors and phenomena that might influence outcomes). Analysis of the simulation results can support computational decision analysis (ex ante evaluation of policy options) and game-theoretic analysis (utility ranking of the alternative courses of actions for an actor).

Such exploratory approach can contribute to improved analysis of complex and uncertain problems by revealing the implications of our knowledge and assumptions, which may include unexpected impacts (Steve Bankes, 1993).

The operational analysis plays an important role in examining and evaluating the considered alternatives for the policy decision based on empirical analysis. The structured decision evaluation of the generated alternative policy options provides a way to use the logic of evaluation methodology and its procedures to evaluate the consequences of alternative and competing policy proposals in terms of each of the selected criteria.

The deliverable is structured as follows:

Chapter 1 introduces multi-criteria decision analysis, the purpose of the technique, the main steps involved and its application for evaluating competing policy options.
Chapter 2 defines the main criteria for evaluation of policy interventions, along with the process and degree of analysis conducted to construct a data format to judge the performance of a policy intervention against each criterion.
Chapter 3 presents the structure of a "common policy appraisal format", the criteria model and the underlying attributes' data formats.

1 Multi-criteria Decision Analysis (MCDA)

1.1 Description and purpose of the technique

Multi-criteria decision analysis (MCDA) has been extensively used to support a wide variety of complex decision problems as a decision-making tool that is particularly useful where decisions involve the achievement of multiple objectives and considering multiple decision makers and stakeholders. With this technique, it is possible to structure and combine multiple assessment criteria to be applied simultaneously for comparison of a number of decision alternatives or choices. See e.g., (Zavadskas et al. 2014, Saaty T.L., 1984, Roy, B., & Bouyssou, D., 1993).

Participation of the decision-makers in the process is a central part of the approach. The method is designed to help decision-makers to assess the performance of different options, reflect the opinions of the involved actors, into a prospective or retrospective framework. Multi-criteria evaluation can be organized with an objective to produce a single synthetic conclusion or to produce conclusions adapted to the preferences and priorities of several actors.

MCDA can be used when several points of view need to be taken into account, when a choice between judgement criteria is difficult or with contradictory criteria (for example, the economic value and environmental impact). Importantly, MCDA is used to highlight the reasoning and subjective views of the different stakeholders. It is usually used to synthesise the opinions expressed, in order to determine the priority structures, to analyse conflictual situations, or to formulate operational advice or recommendations for future activities.

In general, this technique is mainly used in ex ante evaluations of public programs and projects. However, it probably has potential for wider use as a tool in intermediate and ex post evaluations as an aid for making a judgment.

As an evaluation framework, MCDA can also encompass a variety of analytical methodologies, including cost-benefit analysis. It can identify the alternative that, for a given output level, minimises the actual value of costs, or, alternatively, for a given cost, maximises the output level. Results from cost-benefit analysis, scenario or simulation modelling, input-output models, environmental impact assessment, cost effectiveness analysis and damage assessment, energy analysis, or any of the more disciplinary specific tools can each contribute to the quantitative and qualitative information necessary for MCDA.

There two different approaches for combining the set of evaluation criteria for selecting an option to be synthesized or to rank the decision options:

(i) Equally weighted criteria: Each criterion is given equal weight. Thus, a high score on one criterion might offset a lower score on another. Or, alternatively, a threshold score in all criteria may be required for the option to be considered.

(ii) Unequally weighted Criteria: The criteria weights express the importance of each criterion relative to other criteria. Decision-maker's preferences are expressed as weights of relative importance assigned to the evaluation criteria. Deriving the weights is a key step in eliciting the decision-makers preferences. This approach has the advantage of better representing the importance of different criteria. It is still possible, however, that modest strength on several relatively less important criteria can offset a serious flaw on a significant criterion, if scores on each criterion are aggregated.

MCDA provides a framework for communicating the feasible alternatives and the implications of actions. This allows all the involved actors can take part in decision-making and in problem solving, in order to balance their competing objectives. Through explicit treatment of evaluation criteria and negotiation between stakeholders, the technique provides a work process that allows the values and individual opinions of several actors to be taken into consideration, and the processing of functional relations within a complex network, in a quantitative way.

The involvement of experts, the margin of manoeuvre enjoyed by decision-makers and similarities with vote-based methods makes this a suitable tool for a partnership approach. Furthermore, the technique may help to reach a compromise or define a coalition of views, but it does not dictate the individual or collective judgment of the partners.

For instance, in the case of European Union socio-economic programmes, the different levels of partnership (European, national and regional) may be concerned. Each of these levels is legitimate in establishing its own priorities and expressing its own preferences between criteria (EVALSED Sourcebook, 2013). On the issue of the multi-valued nature of environmental policy, Proctor (2000) suggests that: "*Governments now realize that consideration and effective integration of all resource values, whether they are environmental, economic or social, is a necessary first step to achieving and maintaining ecologically sustainable development*" (Proctor, 2000, p. 1).

However, specific problems of implementation may limit the use of MCDA, or require the presence of experts. In addition, this technique is not always used in an interactive way, as it should be, and tends to fix criteria that are, in reality, fluid.

1.2 Main steps involved

Step 1. Definition of the alternative actions to be evaluated

- The set of planned or implemented actions, decision options or the elements on which the comparative evaluation will be made. This the set of "policy options" to be evaluated.

Step 2. Definition of the evaluation criteria

The set of evaluation criteria must be complete, operational, decomposable, non-redundant and minimal. The evaluation criteria summarise and group together diverse characteristic dimensions used to evaluate an action in the form of a statement of objectives and the alternative actions for achieving the objectives. An attribute is used to measure performance in relation to an objective, and a measurement scale is needed for each attribute. Particular attention must be given to the definition of criteria, in order to be as exhaustive as possible and to define the question properly. Unlike the number of decision options to be compared, which can be very large, the number of criteria must not exceed a reasonable limit.

Step 3. Analysis of the impacts of the actions

Both quantitative and qualitative assessments of the impacts of each action can be done in terms of the evaluation criteria and the underlying measurable attributes. These assessments can be quantitative figures (percentage changes) or qualitative descriptors of the different levels of impact. Usually, a Multi-criteria evaluation matrix is constructed with as many columns as there are criteria and as many lines as there are decision options or alternatives to be compared. Each cell represents the evaluation of one alternative for one criterion. MCDA requires an evaluation of all the decision options for all the criteria, but does not require that all the evaluations take the same form. The technique can support a mix of quantitative criteria expressed by indicators, qualitative criteria expressed by descriptors, and intermediate criteria expressed by scores (e.g., a scale 0-10). Table 1 shows an example for a multi-criteria evaluation matrix which can be viewed as an impact scoring matrix. Unequal weights are assigned to criteria (note that the sum of the weights $w_1, \dots, w_n$ equals 1).

Decision Factors	Criterion weight	Economic w_1	Financial w_2	Social w_3	Environmental w_4
Decision option	1	20%	510.5	High impact	4
Decision option	2	35%	312	Low impact	2
Decision option	3	10%	615	Neutral impact	6
⋮	⋮	⋮	⋮	⋮	⋮
Decision option	n	10%	320	High impact	5

Table 1 : Multi-criteria evaluation matrix

The decision options can be compared using one of two variants of the MCDA:

1. *Compensation method*: It consists of attributing a weight to each criterion and then of calculating a global score for each decision option, in the form of a weighted arithmetic average of the scores attributed to that option for the different criteria. This variant is called "compensatory" because the calculation of the weighted average makes it possible to compensate between criteria. For example, a decision option which had a low impact on a criterion could still obtain a good global weighted score if its impact on another criterion were considered excellent.
2. *Outranking method*: This variant is used where the criteria are not all considered commensurable, and therefore no global score can be calculated. The analysis is based on pairwise comparisons of the decision options, (e.g., “Does option 1 outrank option 2 from the point of view of the economic criterion”, “Does option 1 outrank option 2 from the point of view of the environmental criterion” … etc.). These questions can be answered yes or no or be qualified, in which case the notions of a weak preference and a threshold criterion are introduced. The analysis makes all possible comparisons and presents a synthesis in relation to a majority of criteria (case of agreement), without being altogether too bad in relation to the other criteria (case of disagreement)". The analysis could include protection against a favourable judgement for an option that would be disastrous from the point of view of a given criterion, by setting a 'veto threshold' for each criterion. The introduction of a veto threshold strongly differentiates the logic of outranking from the logic of compensation. For example, if there were a veto threshold, a very bad impact on one criterion would make it impossible to consider the option good, even if its impact on another were considered excellent.

From a technical point of view, the compensation variant is easier to implement. Outranking does not always produce clear conclusions, whereas analysis based on compensation it is always conclusive. Usually the compensation method is used unless members of the steering identify a problem which might justify the use of the veto system.

Outranking has the advantage of reflecting the nature of relations between public institutions better, since there is often a correspondence between evaluation criteria and evaluation stakeholders. This probably makes the outranking variant better reflect the collective process of formulating a judgement within the decision makers.

Step 4. Evaluation of the effects of the actions in terms of each of the selected criteria

The process of evaluating impacts could be based on quantitative data by allocating scores to the decision options, or, undertaken more subjectively, by experts or the stakeholders of the evaluation themselves. In reality, the technique usually combines factual and objective elements concerning impacts, with the points of view and preferences of the main partners or members of the evaluation steering group (assessors). The assessors' preferences can be revealed through “direct expression” in the form of a weighting attributed to each criterion or stating their preferences between decision options, or through “classification of profiles”, where the assessors are presented with profiles regarding decision options that reveal preferences between criteria.

Step 5. Aggregation of judgments

Given the presence of many decision makers or stakeholders, for a group decision a decision has to be made and there might be disagreement with respect to what the best option is, but the group choice still has sufficient support from the group, i.e. there is a way of representing the collective preferences of the group. This differs from negotiation in that in such as process each stakeholder might simply abandon the decision process and there is no decision at all, see, e.g., (Kilgour et al. 2010).

A single weighting system for criteria can be deduced, or the evaluation team and steering group can decide to establish average weightings, which has the effect of effacing different points of view among the assessors.

There are three different approaches to the aggregation of judgments:

1. Personal judgments: where the evaluation criteria are not synthesised in any way. Each participant constructs her or his own personal judgment based on the analysis and uses it to argue her or his point of view.

2. Assisting coalition: the different evaluation criteria are ranked. An option will be classified above another one if it has a better score for the majority of criteria (maximum number of allies) and if it has less 'eliminatory scores' compared to the other criteria (minimum number of opponents).

3. Assisting compromise: a weighting of the criteria is negotiated and proposed by the decision makers. The result is a classification of options in terms of their weighted score. The global score for each decision option is calculated by multiplying each elementary score by its weighting and by adding the elementary weighted scores.

The synthesized judgment on the effectiveness of policy interventions would be considered sound and impartial provided that:

(i) the evaluation criteria have been validated by the steering group;
(ii) the conclusions on the impacts of each decision option, as well as the impact scoring matrix summarising them, have been validated;
(iii) the weighting coefficients for criteria, have been established with the assistance of the assessors and the agreement of the steering group.

Experience also shows that the partners are far more willing to accept the conclusions of the report if the evaluation team has recorded their opinions carefully and taken the trouble to take their preferences into account in presenting its conclusions (EVALSED Sourcebook, 2013).

1.3 Multi-criteria Evaluation of policy options

When evaluation of policy options is guided by a decision analysis approach, a relevant MCDA model is required equipped with preference elicitation methods for capturing policy makers' and stakeholders' preferences. Applying MCDA should thus provide the relative global performance of each alternative, and is particularly useful when selecting one out of a finite set of feasible alternatives. Given that we have identified feasible policy options from scenario generation, there are two main tasks remaining in structuring MCDA evaluation models; i) representation of objectives in a structure, commonly a value tree, and ii) the definition of attributes to measure the achievement of objectives (Franco and Montibeller 2010).

Traditionally in MCDA, the decision process starts by structuring the problem as an attribute tree hierarchically ordering the decision makers' aims at different abstraction levels, from fundamental objectives (such as “improved environment”) to lower level attributes (such as “CO_2 emissions reduction”) where the latter contributes to the former in a hierarchical value tree. It is generally assumes that each criterion can be operationalized by a set of measurable attributes allowing for assessing the consequences arising from the implementation of any particular alternative. In the next step, preferential information is elicited. The relative importance of criteria is captured in weights for each criterion at each abstraction level. At the lowest level of the value tree these objectives are translated into attributes, with each one of them evaluating a given characteristic of the decision options (for example, an objective ‘efficiency’ may be measured by the attribute 'operating cost'). The performance of each decision option against each attribute is determined and weights reflecting acceptable trade-offs of performance among objectives are elicited from the decision-makers.

MCDA typically distinguish between a decision maker which has control of some decision variable, and stakeholders, which are affected by the consequences of a decision. This distinction is straightforwardly analogous to the actor representations in the policy model. The simulation is run upon the policy model, whereas the set of objectives and their target values are used for identifying feasible scenarios. Based on the simulation results of an initially large sample of scenarios generated by, e.g., full factorial design or Latin hypercube sampling, unsatisfactory scenarios not realizing the goal vector or being dominated are filtered out, while scenarios deemed efficient and “interesting” according to some predefined decision rule based upon resource constraints and goal compatibility concepts mentioned above are suggested as policy options for further evaluation. Arriving at a set of feasible options which are non-dominated, further discrimination between them will call for taking preference information into account.

Comes et al. (2011) presented the concept of “decision maps”, integrating problem structuring and scenario planning using causal maps with MCDA, but keeps the focus on operational decision making and to a lesser extent on policy making. Each decision map consists of two parts: a causal map and an attribute tree. However, the structure and content of the map allows for efficiently processing information relevant for the decision at hand, proceeding from causes to effects. It also informs the building of an attribute tree, in an ad hoc translation, which enables an assessment of the resulting scenarios with respect to multiple goals and the decision makers' preferences (Comes et al., 2011).

Montibeller and Belton (2006) investigate various ways to use causal maps as the underlying problem structuring tool and extending it with decision evaluation features and/or using the map to effectively inform the decision analysis model in the form of a multi attribute value tree. For simulation, in the SEMPAI[1] framework reported in Hansson et al. (2011), simulation results of flooding models are combined with MCDA where multiple stakeholders are present.

As a simple example, consider a local government wanting to reduce CO2 emissions in order to contribute to climate targets and improve central town air quality. The upfront objectives are focused on the environmental issue. One policy option is to increase the subsidy of bus tickets, where the underlying hypothesis in terms of causality is that higher subsidies (lower ticket prices) will cause more people to take the bus instead of cars when going into town, thereby reducing emissions also enabling for more citizens to access the central town area increasing town commerce, which is the main concern of an influential stakeholder group. However, there is also a positive causal relationship between car traffic and town commerce, increasing the

[1] Simulation and Evaluation with Multiple Perspectives and Agents Integrated

complexity of the policy option making. In the problem analysis, it becomes clear that for many citizens it is not the bus ticket price that cause them to avoid taking the bus, but rather the frequency of buses. Further, reducing emissions can also be done by increasing the proportion of gas driven buses. For such example, see Figure 3 showing a value tree model based upon the causal mapping simulation in Figure 2.

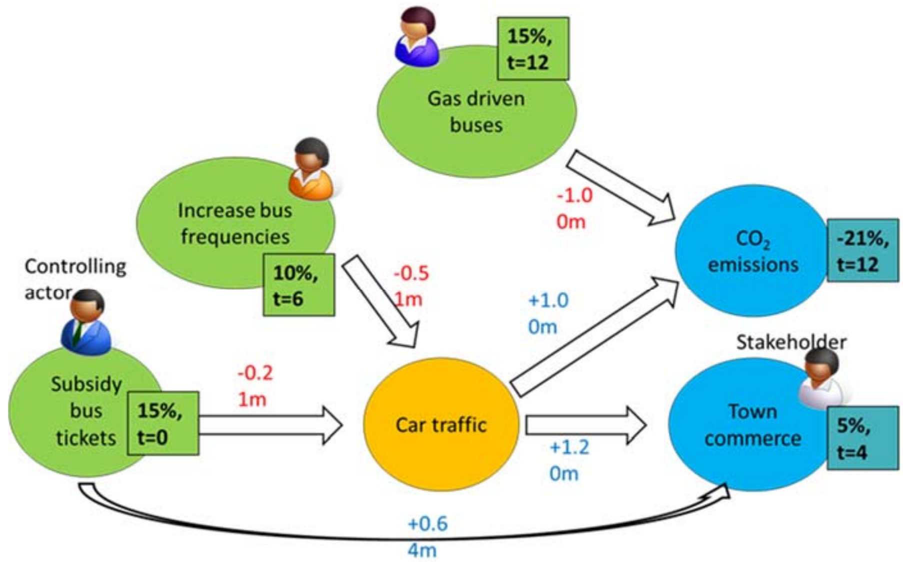

Figure 2 : Example of a generated scenario reaching both the emission and the town commerce target. The scenario consists of increasing the subsidy with 15% at the initial time-step, increasing bus frequencies at the sixth, and increasing the proportion of gas driven buses at time-step (t=12).

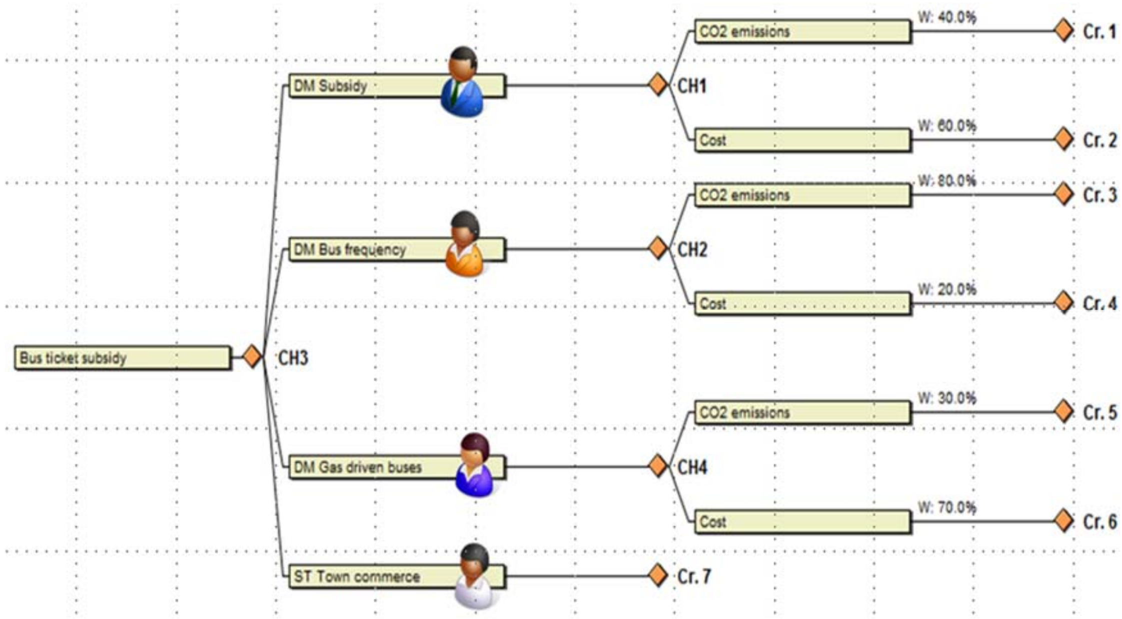

Figure 3 : Value tree counterpart of causal map in Fig. 3 in accordance to the SEMPAI framework. In this example, DM Subsidy's weight for CO2 emissions is 40% so for simplicity precise numerical weights are used here.

Elicitation or anticipation of decision maker and stakeholder preferences is done or assessed in the form of utility statements for each actor or actor group regarding each specific policy option and criteria weights. Actors give the different outcomes a ranking order if they are unsure of their preferred choice and methods for computational decision analysis with imprecise information is promoted in order to support such statements, see (Larsson et al. 2005; Danielson et al. 2007). The result is a preference assessment for each actor and the value tree is constructed with actors as the lowest level of the tree, see Fig. 3 for the value tree corresponding to the causal map of Fig. 2 where each actor has a trade-off expressed as criteria weights between cost and their benefits of concern.

With respect to the MCDA model in Figure 3, we can consider the following simple preference scheme using value functions V(S) of the scenarios for each criterion delimiting the decision evaluation to the reference scenario called S0 and the generated scenario shown in Fig. 3 called S1. Here the value function V indicates the actors' preferences as weights for the two scenarios (decision options), from the point of view of each criterion, (e.g., DM subsidy prefers S1 with respect to "CO2 emissions" criterion).

Decision maker	CO2 emissions		Cost	
	V(S0)	V(S1)	V(S0)	V(S1)
DM Subsidy	0	1	1	0
DM Bus frequency	0	1	1	0
DM Gas driven buses	0	1	1	0

Stakeholder	Town commerce	
	V(S0)	V(S1)
ST Town commerce	0	1

Table 2 : Preference scheme

Decision evaluation is then done actor-wise. Each actor (decision maker or stakeholder) will receive a utility value for each scenario and the choice of selecting or discarding one scenario will be based upon its total utility through aggregating the actors' utility values (see Fig. 5) together with group decision admissibility concepts such as maximum disagreement thresholds or minimum consensus thresholds (Fasth et al. 2013).

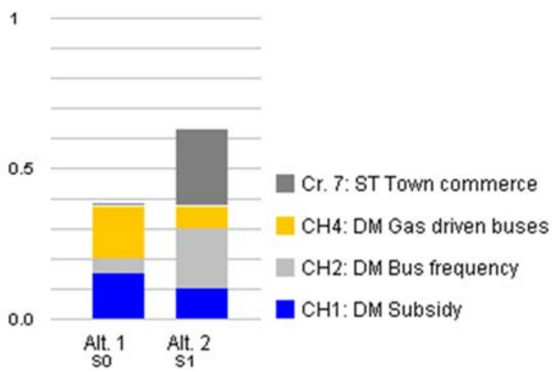

Figure 4 : Comparison of scenarios generated by adding part-worth utilities for each actor. There is disagreement since DM Subsidy and DM Gas driven buses consider S1 to be inferior to S0 while the other actors have differing preferences.

2 Evaluation Criteria

The European commission's better regulation guidelines and the corresponding toolbox[2] regarding the evaluations and fitness checks asserted that all evaluations of policy interventions must assess the evaluation criteria of **effectiveness, efficiency, coherence, relevance** and **added value** of the intervention, and additional criteria beyond these five can be added. This chapter discusses each of these key evaluation criteria along with the process and degree of analysis conducted to construct a data format to judge the performance of a policy intervention against each criterion.

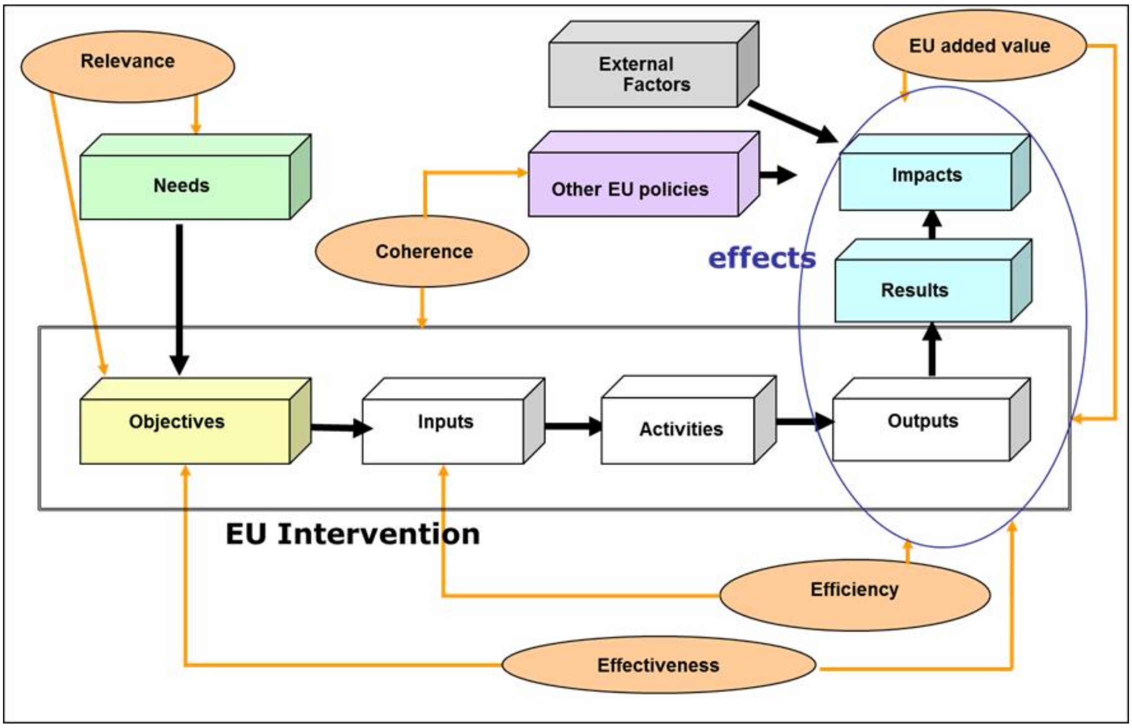

Figure 5 : The simplified EU intervention logic and the 5 key evaluation criteria.

2.1 Effectiveness

Effectiveness analysis considers how successful the policy intervention is in achieving or progressing towards its objectives. The evaluation should form an opinion on the role of the action in delivering the observed changes (from simulation results). The analysis should also try to identify if any unexpected or unintended effects have occurred.
Examples of effectiveness questions:
• To what extent are the objectives achieved?
• What are the (quantitative and qualitative) effects of the intervention?
• To what extent do the observed effects correspond to the objectives?
• To what extent can these changes/effects be credited to the intervention?
• What factors influenced the achievements observed?
• To what extent did different factors influence the achievements observed?

A policy option was defined in our policy modelling approach as a scenario of change triggered by a policy instrument (mostly in combination with others). Policy instruments can vary

[2] http://ec.europa.eu/smart-regulation/guidelines/index_en.htm

significantly in their effectiveness in different countries. Therefore, it is important to identify key factors determining the effectiveness of the instrument as well as barriers which can explain failure of the same instrument.

The effectiveness of policy instruments can be analysed in both qualitative and quantitative ways. For example, considering the improvement of energy efficiency and the reduction of CO_2 emissions, to be the most important attributes for evaluating effectiveness of the energy transition and climate action policies or programs. Each policy instrument's effectiveness can be analysed using the amount (or percentage of baseline) of energy use or emissions saved as a result of the policy. In cases of insufficient quantitative data, absence of baselines or difficulties with comparing the numerical values of emission reductions as their temporal and spatial scale differs as well as the total emission coverage we can rely on the available data and experts' judgements. Alternatively, qualitative grades (e.g., "High", "Medium" and "Low") can be assigned to policy instruments for their performance in reducing energy use and emissions due to limitations of availability and reliability of data.

2.2 Efficiency

Efficiency considers the relationship between the resources used by an intervention and the changes generated by the intervention. The financial impact variables in the policy model should reflect the costs on governmental departments, businesses and citizens (when applicable).

Efficiency analysis includes analysis of ancillary costs, administrative and regulatory burden. Findings should pin-point areas where there is potential to reduce inefficiencies and simplify the intervention. The efforts to support and perform an intervention include: staff, time and/or money spent, fixed costs, running costs, etc. These costs can be associated to different aspects of an intervention and judged against the benefits achieved. It is important to note that efficiency analysis should always assess both the costs and benefits of the intervention. Efficiency (cost-effectiveness) as viewed from a societal perspective should take into account not only investment and capital costs as well as implementation costs, but it should include the financial benefits (i.e. avoided costs) as a result of the policy intervention.

Examples of efficiency questions:
- To what extent are the costs involved justified, given the changes/effects which have been achieved?
- To what extent are the costs proportionate to the benefits achieved?
- How affordable were the costs borne by different actor groups, given the benefits they received?
- What are the costs on governmental departments and businesses, including small businesses?
- For EU policies, if there are significant differences in costs (or benefits) between Member States, what is causing them?

Like effectiveness, efficiency can be evaluated in a quantitative and a qualitative way. Calculations of quantitative data include costs per unit of the achieved goal (e.g., cost of reduced emissions in EUR/t CO_2 eq.). All financial values need to be deflated and benefits of avoided costs need to be included. The qualitative grades can be based again on the available quantitative figures and experts' judgment.

2.3 Relevance

Relevance looks at the relationship between the needs and problems in society and the objectives of the intervention.

For example, the wrong "problem drivers" may have been identified during the impact assessment; incorrect assumptions may have been made about the cause and effect relationships.

Examples of relevance questions
• To what extent have the objectives proven to be appropriate for the intervention?
• How well do the objectives correspond to the needs and problems?
• How well adapted is the intervention to subsequent technological or scientific advances?
• How relevant is the intervention to citizens?

2.4 Coherence

The evaluation of coherence involves looking at how well or not different actions work together. Checking "internal" coherence means looking at how the various internal components of an intervention operate together to achieve its objectives (multiple policy instruments). Similar checks can be conducted in relation to (a limited number of) other ("external") interventions, at different levels: for example, between interventions within the same policy field. At its widest, external coherence can look at compliance with international agreements/declarations.

Examples of coherence questions
• To what extent is this intervention coherent with other interventions which have similar objectives?
• To what extent is the intervention internally coherent?
• To what extent is the intervention coherent with wider EU policy?
• To what extent is the intervention coherent with international obligations?

2.5 Added value

Added value looks for changes which it can reasonably be argued are due to the intervention, rather than any other factors. In many ways, the evaluation of the added value brings together the findings of other criteria (e.g., relevance and efficiency), presenting the arguments on causality and drawing conclusions, based on the evidence at hand, about the performance of the intervention. For example, assessment of the added value of an EU intervention requires consideration of the change and improvements which are caused by the EU rather than another party taking action, e.g., coordination gains, legal certainty, greater effectiveness, complementarities … etc. In all cases, measurement is a challenge and the final judgement on whether expected added value would justify an intervention is ultimately the result of a political process.

Examples of added value questions
• What is the additional value resulting from the EU intervention(s), compared to what could be achieved by Member States at national and/or regional levels?

• To what extent do the issues addressed by the intervention continue to require action?
• What would be the most likely consequences of stopping or withdrawing the existing intervention?

2.6 Other evaluation criteria

There are also several further evaluation criteria which it may be appropriate to consider, depending on the type of intervention and the timing of the evaluation. The most common additional criteria evaluated by the Commission are shown below.

Utility: To what extent do the changes/effects of an intervention satisfy (or not) stakeholders' needs?

Complementarity: To what extent does an intervention support and usefully supplement other policies.

Equity: how fairly are the different effects distributed across the different stakeholders / regions? / genders? / Social groups?

Sustainability: How likely are the effects to last after the intervention ends? It can be important to test this expectation for interventions which have a finite duration, such as particular programmes.

Acceptability: To what extent can we observe changes in the perception of the intervention (positive or negative) by the targeted stakeholders and/or by the general public?

3 Common Policy Appraisal Format

The following figure illustrates the structure of the common policy appraisal format, to assess the global performance of policy options, as a multi-attribute value tree with two abstraction levels for the fundamental evaluation criteria and the underlying attributes. The data formats of the attributes are based on the simulation results (Policy options and Impact assessments).

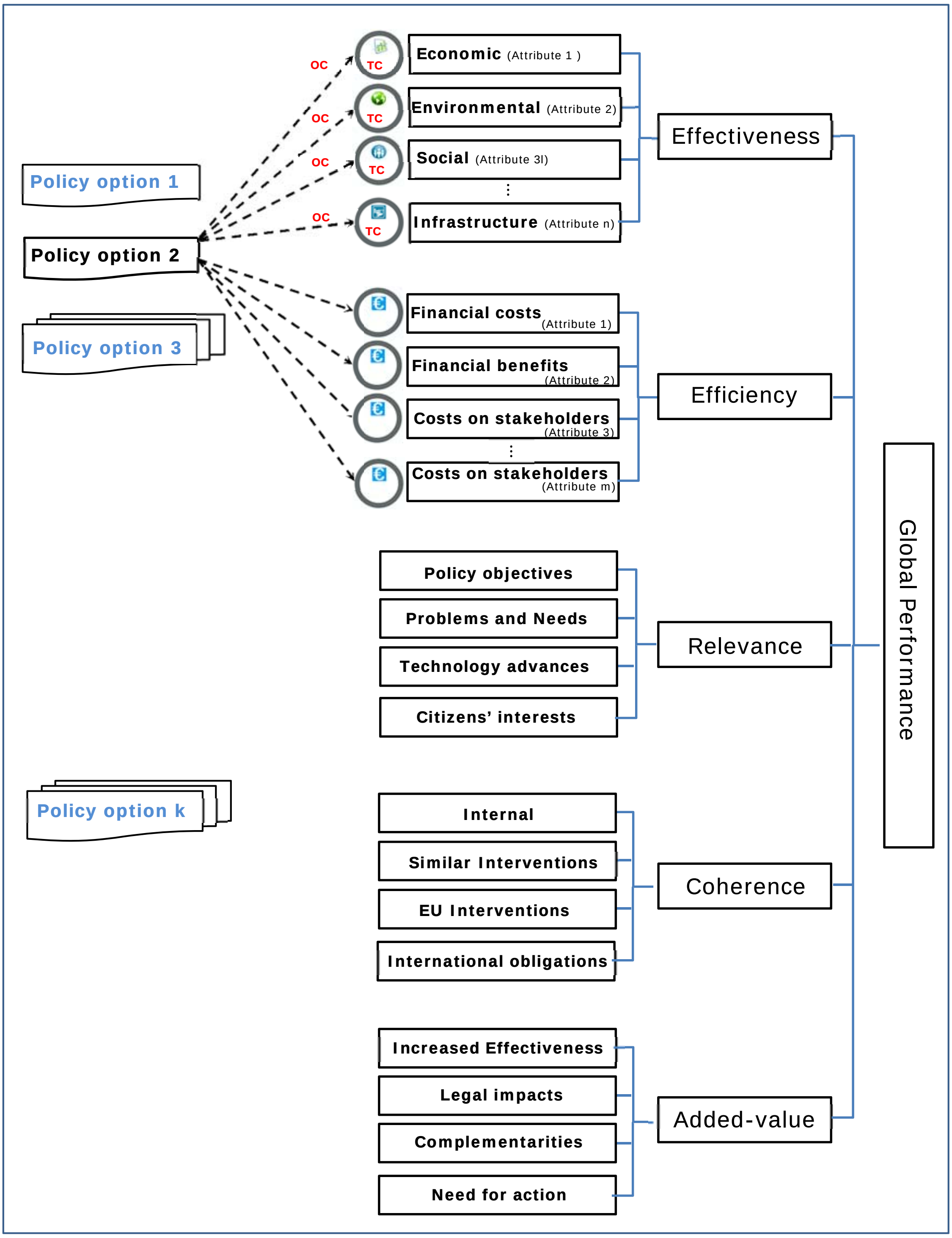

Figure 6 : Common Policy Appraisal Format.

The figure shows ***k*** policy options to be evaluated. For each policy option, the impact assessment results for each policy option provide the observed changes on the goal variables as compared to the targeted changes. Note that the appraisal is done from the perspective of one actor, who is requested to express goal preferences and evaluations of policy options. Support methods and a process for this is subject of a coming Deliverable 6.4.

Each of the attributes underlying the “effectiveness” evaluation criterion reflects how successful a policy option is in achieving the policy goals by comparing the observed change (OC) to the targeted change (TC) for each impact variable or for impact variables within the same category (e.g., economic, social, environmental … etc.). Unintended or unexpended effects of a policy action can also be included by adding the corresponding impact variable and attributes.

The attributes underlying the “efficiency” criterion are concerned with the financial impacts of a policy option, whether financial costs or benefits (cost savings). The evaluation of the impact of the action is expressed as the costs involved on governmental departments given the achieved changes. In addition costs on businesses (including small businesses), or different stakeholder groups are expressed in terms of different attributes. The efficiency evaluations can be made in a quantitative or a qualitative way.

The attributes underlying the “Relevance” criterion are using qualitative descriptors (or value scales) to assess the relevance of a policy option to the policy objectives, to the problems/needs, to the technological advances and to the citizens.

The attributes underlying the “Coherence” criterion are using qualitative descriptors or value scales to assess the coherence of a policy option: (i) internally if it includes multiple policy instruments, (ii) with other similar interventions, (iii) with EU policy interventions and (iv) with international obligations.

Finally the attributes underlying the “added-value” criterion are using qualitative descriptors or value scales to assess the added value of the policy option in terms of legal impacts, increased effectiveness, complementarities and the need for continuing this intervention.

4 Conclusion

As a conclusion, the proposed common policy appraisal format provides a basis for further computational decision analysis for evaluation of policy options beyond impact assessment only, based upon a set of common criteria being linked to a set of context dependent attributes enabling for use of preference statements when comparing policy options. This format is used to guide the user into considering each of the five primary evaluation criteria aspects.

Identifying the actor groups enables appraising a policy from the perspective of multiple decision makers and stakeholders with different priorities and preferences, as well as for systematic tools for eliciting preferences from stakeholders and decision makers. This work is subject to forthcoming deliverable D6.4.

5 References

Axelrod, R. (1976). The cognitive mapping approach to decision making. In Axelrod, R. (ed.), Structure of Decision, Princeton University Press, Princeton NJ, pp. 221-250.

Bankes, S. (1993). Exploratory modeling for policy analysis. Operations Research, 41(3), 435-449.

Comes, T., Hiete, M., Wijngaards, N., and Schultmann, F. (2011). Decision maps: a frame-work for multi-criteria decision support under severe uncertainty, Decision Support Systems 52: 108-118.

Danielson, M., Ekenberg, L., Idefeldt, J., and Larsson, A. (2007). Using a software tool for public decision analysis: the case of Nacka municipality, Decision Analysis 4(2):76-90.

EVALSED Sourcebook – Evaluation Method and techniques. The resource for the evaluation of Socio-Economic Development: European Commission, 2013. (online- retrieved on 1/9/2015)

Fasth, T., Kalinina, M., and Larsson, A. (2013). Admissibility concepts for group portfolio decision analysis, Proceedings of GDN 2013.

Franco, L. A. and Montibeller, G. (2010). Facilitated modelling in operational research. European Journal of Operational Research 205: 489-500.

Hansson, K., Larsson, A., Danielson, M., and Ekenberg, L. (2011). Coping with complex environmental and societal flood risk management strategies: an integrated multi-criteria framework, Sustainability 3(9):1357-1380.

Hamilton, A. (2010). Policy formulation – critique, analysis and strategic implications, in New Voices: Essays on the Policy Process, JSGS Student Working Paper Series, Johnson-Shoyama Graduate School of Public Policy.

Kilgour, D. M., Chen, Y., and Hipel, K. W. (2010). Multiple criteria approaches to group decision and negotiation, in Ehrgott et al. (eds.), Trends in Multiple Criteria Decision Analysis, Springer.

Koomey, J., Rosenfeld, A. H., and Gadgil, A. K. 1990. Conservation Screening Curves to Compare Efficiency Investments to Power Plants. Energy Policy 18 (8), 774-782.

Larsson, A., Johansson, J., Ekenberg, L., and Danielson, M. (2005). Decision analysis with multiple objectives in a framework for evaluating imprecision, International Journal of Un-certainty, Fuzziness and Knowledge-Based Systems 13:495-510.

Montibeller ,G. and Belton, V. (2006). Causal maps and the evaluation of decision options – a review. Journal of the Operational Research Society 57: 779-791.

Proctor, W. (2000). Can Multi-criteria Analysis Aid the Process of Environmental Policymaking? International Society for Ecological Economics 2000 Conference: Operationalising Ecological Economics, The Australian National University, Canberra.

Roy, B., & Bouyssou, D. (1993). Aide multicritère à la décision: méthodes et cas. Paris: Economica.

Saaty T.L. (1984), Décider face la complexité, Paris: Entreprise Moderne d'Edition,

Schärlig A. (1990, 2me d.), Décider sur plusieurs critères, panorama de l'aide la décision multicritère, Lausanne: Presses polytechniques et universitaires romandes,

Turnpenny, J., Radaelli, C. M., Jordan, A., and Jacob, K. (2009). The policy and politics of policy appraisal: emerging trends and new directions. Journal of European Public Policy 16(4):640-653.

U.S. General Accounting Office, Program evaluation and Methodology division. Prospective Evaluation Methods. The Prospective Evaluation Synthesis. GAO/PEMD-10.1.10, Washington D.C.: November 1990.

Zavadskas, E. K., Turskis, Z., Kildiene, S. (2014). State of art surveys of overviews on MCDM/MADM methods, Technological and Economic Development of Economy 20(1):165-179.