

Romualdas GINEVICIUS, PhD

romualdas.ginevicius@mruni.eu

Mykolas Romeris University, Vilnius, Lithuania

Dainora GEDVILAITE, PhD

d.gedvilaite@ekf.viko.lt

Vilniaus kolegija/University of Applied Sciences, Vilnius, Lithuania

Yurij BILAN, PhD (corresponding author)

y.bilan@csr-pub.eu

Széchenyi István University, Gyor, Hungary

Pan-European University, Bratislava, Slovakia

Viktorija STASYTYTE, PhD

viktorija.stasytyte@vilniustech.lt

Vilnius Gediminas Technical University, Vilnius, Lithuania

Andrius STASIUKYNAS, PhD

andrius.stasiukynas@vilniustech.lt

Vilnius Gediminas Technical University, Vilnius, Lithuania

Comprehensive Assessment of the Reliability of Verbal Information

Abstract. *Multi-criteria methods commonly evaluate indicators based on two dimensions – indicator significance and size. As the former is typically evaluated by experts, it is necessary to assess expert opinion consistency adequately. Accordingly, this study aims to develop a more comprehensive methodology for measuring expert opinion consistency in multi-criteria assessment. The widely used methodology, which is based on Kendall’s criterion W or Pearson’s criterion χ^2 , provides only a very rough picture. This is for two reasons. First, expert opinion consistency is measured not on the basis of initial information, but on the basis of transformed values – ranks. Second, expert opinions are measured in a non-complex way, i.e. based only on differentiation of the aggregate estimate of an indicator’s significance provided by all experts, while differentiation of an individual expert’s evaluations across all indicators is not considered. The methodology proposed in this article eliminates both above-mentioned shortcomings by combining horizontal and vertical differentiation of the original expert evaluations into a single consistency indicator.*

Keywords: multi-criteria assessment, expert opinion evaluation, expert opinion consistency.

JEL Classification: C02, C44, C46, C52, D81.

Received: 8 August 2026	Revised: 16 September 2026	Accepted: 21 September 2026
-------------------------	----------------------------	-----------------------------

1. Introduction

1.1 Expert evaluation and criterion weighting

Experts have an important role to play in analysing and evaluating different economic, social and managerial processes as many of the indicators that describe the above-mentioned processes are difficult to formalise. This situation is mainly faced when assessing the actual state of the development of particular socio-economic systems (SEs) by employing multi-criteria methods (Hwang & Yoon, 1981; Gedvilaite, 2019; Oželiene, 2019; Mateiaş & Kanungo, 2025) because SEs are usually large, complex and manifest in many different aspects. The indicators describing these aspects are ambiguous in their impact on a phenomenon under consideration, which is why their ultimate value is estimated as the product of the initial value and the significance of an indicator (Hwang & Yoon, 1981; Hwang & Lin, 1987; MacCrimmon, 1968). The determination of indicator significance is therefore a central issue in multi-criteria evaluation. Ginevičius et al. (2023) show that the ranking of indicator importance can be simplified by using the transitivity of expert judgments. Their results also underline that the credibility of final indicator weights depends on the quality and consistency of the expert information used in their determination.

The value of some indicators can be estimated accurately, based on statistical data, other information sources, calculations, etc., while the significance of these indicators can be evaluated only by experts.

In practical multi-criteria applications, the procedure used to determine criterion significance may substantially affect the final ordering of alternatives. Łopatka et al. (2023) demonstrate this issue in a contractor-selection problem: different criterion-weight configurations produced different rankings of the evaluated alternatives. Therefore, before expert-derived weights are incorporated into a decision model, the reliability of the underlying expert judgments should be assessed.

It is stipulated that the number of experts should be not less than the number of indicators plus 1, i.e. $m + 1$ (Šaparauskas, 2004). In practice, this number is usually higher because this way the probability of anomalies or clear subjectivity of the result is reduced (Bechelev & Gurvich, 1974). Classical test theory proposes that reliability of expert evaluations and the number of experts are related by a rapidly fading nonlinear relationship (Libby & Blashfield, 1978) (Figure 1).

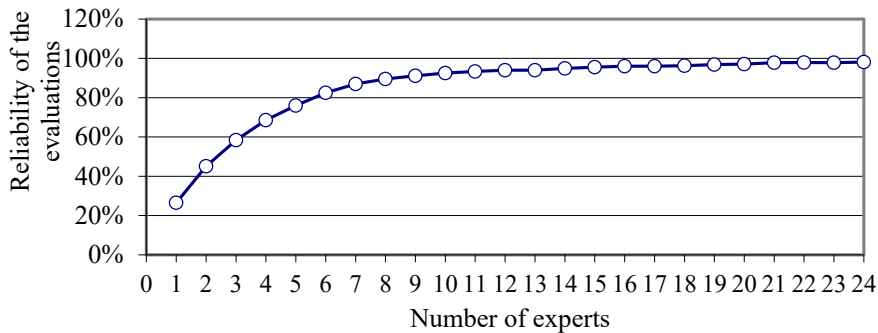


Figure 1. Dependence of the standard deviation of expert evaluations on the number of experts

Source: Libby & Blashfield, 1978.

Figure 1 indicates that when the number of experts is higher than 11, the accuracy of evaluations exceeds 95 percent.

1.2 Expert opinion consistency and methodological limitations

As expert opinions are subjective, it is necessary to assess whether they are consistent. The wider literature on multi-criteria assessment also confirms that the selection and aggregation of indicators must be transparent. Ginevičius et al. (2022) applied qualimetric methods to evaluate complex social processes through integrated indicators. Their study illustrates that the validity of a composite assessment depends on a clear relationship between primary evaluations and the final aggregate result. The literature proposes a variety of methods for measuring consistency. The application of multi-criteria assessment also requires a transparent procedure for selecting, structuring, and combining evaluation criteria. Mubarak et al. (2024) developed a multi-criteria assessment methodology for evaluating the remanufacturing of conventional grinding machines into CNC machine tools on the basis of cost, time, accuracy, reliability, processing-shape complexity, and ergonomics. Although their study concerns an engineering application, it demonstrates that the credibility of a multi-criteria result depends on the quality of the information used to assess the relative importance of the selected criteria. Some of them are widely applied (Hwang & Yoon, 1981; MacCrimmon, 1968; Churchman et al., 1954; Klee, 1971; Yoon, 1980; Cheung & Kuen, 1991; Zagorskas et al., 2007; Behm, 2005; Fang et al., 2004; Kartam et al., 2000; Ustinovičius, 2003; Завадскас, 1987; Евланов, 1984; Srinivasan & Shocker, 1973; Yoon, 1980; Keshavarz et al. 2015; Podvezko, 2005; Ubartė, 2017; Volkov, 2018; Zavadskas & Kaklauskas, 1996), while others are less common (Triantaphyllou, 2000; Opricovic & Tzeng, 2004; Brans et al., 1984; Zavadskas et al., 2008; Roy, 1968; Dinca et al., 2025; Voda et al., 2025). These methods operate at different stages of the multi-criteria decision-making procedure, including the determination of criterion weights, the aggregation of evaluations, and the ranking of alternatives. Stankevičienė et al. (2019) used the

MULTIMOORA method to evaluate the technology-transfer performance of higher education institutions, while Balcerzak (2016) applied MULTIMOORA to assess the quality of human capital in European Union countries. These applications demonstrate the importance of establishing a reliable basis for the evaluation data before they are synthesised into a final multi-criteria result. This requirement is especially important where expert evaluations are used to establish the relationships between goals, expected outcomes, and performance indicators. Kostiukevych et al. (2020) combined expert evaluation with decomposition and synthesis to assess regional economic development projects. In such applications, insufficient consistency of expert judgments can distort the prioritisation of indicators and weaken the validity of subsequent managerial decisions. One of the most common methods is based on Kendall's coefficient W (when the number of indicators is less than or equal to 7) or χ^2 (when the number of indicators is higher than 7) (Kendall, 1975; Kendall, 1948). Nevertheless, the practice of the application of the above-mentioned methods shows that at best they provide only a rough picture. The need for methodological robustness is further confirmed by studies that compare results obtained through several multi-criteria models. Štreimikienė (2023) used different multi-criteria decision-making models to compare and rank waste-management performance in the Baltic States. Such comparisons show that the final ranking may depend on the method used to process the same evaluation information, reinforcing the importance of examining the reliability of initial expert assessments. The treatment of evaluation information is important not only when measuring agreement but also when constructing the final multi-criteria result. Baležentis (2022) proposed the PROMETHEE-KerCA method, which applies different aggregation rules to develop comprehensive utility scores in policy modelling. This confirms that methodological choices concerning the processing and aggregation of evaluative information can materially influence the final outcomes of multi-criteria analysis. An additional methodological challenge concerns the reduction of multidimensional information into a single integrated indicator. Ivanová and Masárová (2016) used multi-criteria evaluation to assess the innovation performance of Slovak regions, combining several indicators into an aggregate measure. This confirms that the quality of the final result is conditional on the adequacy of the initial values and the procedure used to combine them. The role of consistency is also recognised in methods based on pairwise comparisons. Khodaparasti (2015) applied the analytic hierarchy process to rank marketing-mix elements, using a procedure in which the consistency of comparative judgments is relevant to the credibility of the resulting priorities. This provides an additional argument for improving the assessment of expert opinion consistency in multi-criteria evaluation. Taken together, these studies indicate a need to examine more closely how expert-consistency measures process the original evaluations and why rank-based procedures may provide only approximate results. The present study therefore focuses on improving the adequacy of expert opinion consistency measurement by retaining the information contained in the initial evaluations and considering both horizontal and vertical differentiation.

2. Analysis of the expert opinion consistency measuring methodology based on Kendall's criteria

Expert evaluation of the significance of particular indicators representing the phenomenon under consideration is based on matrix $\mathbf{E} = \|e_{ij}\|$, which incorporates expert evaluations of the significance of each indicator (e_{ij} – j th expert's ranking of the i th indicator, $i = \overline{1, m}$, $j = \overline{1, r}$); here m – the number of indicators; r – the number of experts (Figure 2):

$$\mathbf{E} = \begin{matrix} & e_{11} & e_{12} & e_{13} & \cdots & e_{1j} & \cdots & e_{1r} \\ & e_{21} & e_{22} & e_{23} & \cdots & e_{2j} & \cdots & e_{2r} \\ & e_{31} & e_{32} & e_{33} & \cdots & e_{3j} & \cdots & e_{3r} \\ & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ & e_{i1} & e_{i2} & e_{i3} & \cdots & e_{ij} & \cdots & e_{ir} \\ & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ & e_{m1} & e_{m2} & e_{m3} & \cdots & e_{mj} & \cdots & e_{mr} \end{matrix}$$

Figure 2. Matrix of the expert rankings of the significance of particular indicator

Source: Kendall, 1975.

The evaluations in matrix $\mathbf{E}$ are provided in points on the selected scale. A 10-point scale is commonly selected as most obvious. In general, it does not matter which particular scale is selected as in the second stage of consistency measuring, points are converted into ranks or positions, with most significant indicator being ranked 1, and least significant indicator being ranked m . The result of such transformation of the initial data is matrix $\mathbf{R} = \|r_{ij}\|$, here r_{ij} – the rank awarded by the j th expert to the i th indicator (Figure 3).

$$\mathbf{R} = \begin{matrix} & r_{11} & r_{12} & r_{13} & \cdots & r_{1j} & \cdots & r_{1r} \\ & r_{21} & r_{22} & r_{23} & \cdots & r_{2j} & \cdots & r_{2r} \\ & r_{31} & r_{32} & r_{33} & \cdots & r_{3j} & \cdots & r_{3r} \\ & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ & r_{i1} & r_{i2} & r_{i3} & \cdots & r_{ij} & \cdots & r_{ir} \\ & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ & r_{m1} & r_{m2} & r_{m3} & \cdots & r_{mj} & \cdots & r_{mr} \end{matrix}$$

Figure 3. The initial matrix of the expert rankings of the significance of particular indicator

Source: Kendall, 1975.

Depending on the number of indicators, the degree of expert opinion consistency is measured by either Kendall's coefficient of concordance W or Pearson's correlation coefficient χ^2 (Kendal, 1975):

$$W = \frac{S_f}{S_{\max}}$$

$$\chi_f^2 = \frac{12S_f}{rm(m+1)}.$$

Value S_f is estimated as the sum of the squares of the indicator rank total deviation from the mean of all ranks in matrix **R**:

$$S_f = \sum_{i=1}^n |\bar{r} - \sum_{j=1}^r r_{tj}|.$$

Value $\bar{r}$ is estimated as follows:

$$\bar{r} = \frac{\sum_{j=1}^r r_{tj}}{m}.$$

The maximum possible deviation $S_{\max}$ is estimated as follows:

$$S_{\max} = \frac{r^2 m(m^2 - 1)}{12}.$$

The question arises as to the adequacy of this methodology, i.e. how accurately it measures the degree of expert opinion consistency. Let us analyse an example. Let us presume that the phenomenon under consideration is characterised by 5 indicators, the significance of which was evaluated by 6 experts. The results of the expert evaluation are presented in Table 1.

Table 1. Results of the Expert Evaluation of the Significance of the Indicators Representing the Phenomenon under Consideration

Indicators	Experts						$\sum_{j=1}^r e_{ij}$	$\bar{e}_i$	$\sum_{i=1}^m e_i - \sum_{j=1}^r e_{ij} $
	1	2	3	4	5	6			
1	10	5	9	6	10	6	46	7.67	12
2	8	3	7	3	7	4	32	5.33	12
3	9	4	8	4	8	5	38	6.33	12
4	7	2	6	2	6	3	26	4.33	12
5	5	1	5	1	5	2	19	3.17	11
$\sum_{j=1}^m e_j$	39	15	35	16	36	20	161	–	59
$\bar{e}_j$	7.8	3	7	3.2	7.2	4	–		
$\sum_{j=1}^m e_i - \bar{e}_j $	7.2	6	6	7.2	7.2	6	39.6		

Source: Authors' own creation.

The expert evaluations in Table 1 suggest that they are either inconsistent or the degree of consistency is very low. Based on the current methodology, the values of the indicators are converted into ranks (Table 2).

Table 2. Expression of the Initial Information in Ranks

Indicators	Experts						Total
	1	2	3	4	5	6	
1	1	1	1	1	1	1	6
2	3	3	3	3	3	3	18
3	2	2	2	2	2	2	12
4	4	4	4	4	4	4	24
5	5	5	5	5	5	5	30
Total							90

Source: Authors' own creation.

Based on formulas (1)–(5), $S_f = 360$, $S_{\max} = 360$, and $W = 1.0$, i.e. the expert opinions seem to be ideally consistent but, actually, this result clearly contradicts the initial opinions expressed by the experts.

The first shortcoming is that the expert opinion consistency is measured not on the basis of the initial information (Table 1), but on the basis of the transformed values (Table 2). As a result, there is a shift from a more accurate ranking system to a less accurate one (in our example, from 10 to 5-point system). Hence, we have the situation where, as the number of indicators is decreasing but the ranking system remains unchanged, the gap between the two systems is widening, which leads to the expert opinion levelling.

Another drawback of the methodology is that expert opinions are measured in a non-complex way, i.e. based only on differentiation of the aggregate estimates of an indicator's significance (horizontal evaluation), while differentiation of the same expert's ranks (vertical evaluation) **is not considered**. Only a combination of horizontal and vertical indicator significance evaluations into a single aggregate size can provide an adequate result.

3. The methodology proposed for expert opinion consistency measuring

The methodology proposed in this article is based on the initial premise that the degree of expert opinion consistency depends exclusively on differentiation of the opinions of individual experts, expressed in terms of initial estimates rather than ranks. Table 1 shows that both horizontal and vertical differentiations need to be considered. The former reflects the differences in evaluations of the significance of the same indicator provided by all experts. The more equal are expert evaluations, the greater is differentiation of the aggregate evaluation of an indicator's significance, and the higher is degree of consistency. Vertical differentiation reflects

the differences in evaluations of all indicators provided by the same expert. The higher are the evaluations, the greater is differentiation of the aggregate evaluation of an indicator's significance, and the higher is degree of consistency. Thus, there exists an inverse relationship between horizontal and vertical differentiation. For complex measuring of expert opinion consistency, both of the above-mentioned aspects need to be appropriately combined into a single aggregate size. In this case, the model for expert opinion consistency indicator development would look as follows (Figure 4):

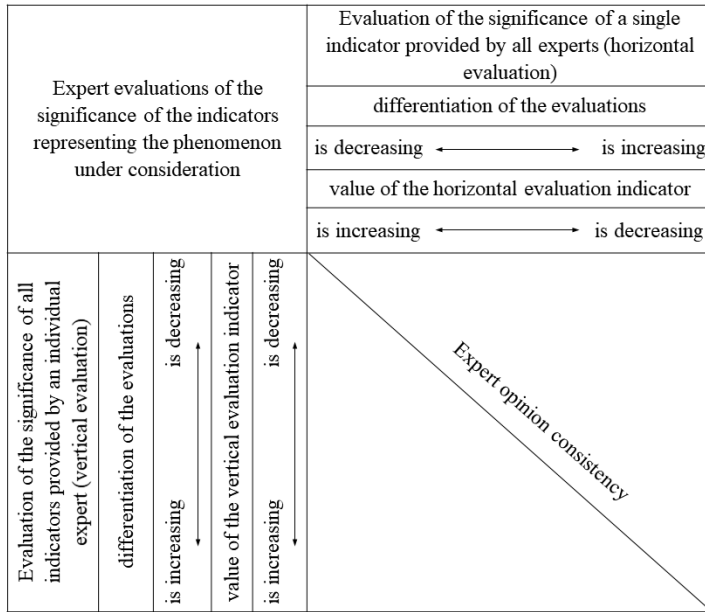


Figure 4. The model for expert opinion consistency indicator development

Source: Authors' own creation.

Figure 4 indicates that, firstly, a high degree of expert opinion consistency is obtained when the equality of expert evaluations concerning the significance of a particular indicator is high (horizontal evaluation) as well as differentiation of the evaluations of all indicators provided by an individual expert (vertical evaluation); secondly, when the equality of horizontal evaluation is increasing, vertical differentiation is also rising.

For the sake of clarity, the methodology proposed is illustrated with two typical examples. In one case, expert opinions are quite unanimous, while in the second case they are different (Tables 3 and 4). The first case is characterized by high vertical differentiation and low horizontal differentiation, while the second case is characterized by low vertical differentiation and high horizontal differentiation.

Table 3. The Results of the Expert Evaluation Concerning the Significance of the Indicators Representing the Phenomenon under Consideration

Indicator	Experts						$\sum_{j=1}^r e_{ij}^h$	$\bar{e}_i^h$	$\sum_{j=1}^r \bar{e}_i^h - \sum_{j=1}^r e_{ij}^h $
	1	2	3	4	5	6			
1	10	8	9	9	10	8	54	9	4.0
2	7	6	5	7	7	6	38	6.33	4.0
3	5	4	5	5	4	5	28	4.67	2.7
4	3	2	4	3	4	3	19	3.17	3.34
5	1	2	1	1	2	1	8	1.33	2.66
$\sum_{i=1}^m e_{ij}$	26	22	24	25	27	23	147	–	16.7
$\bar{e}_j^v$	5.2	4.4	4.8	5.0	5.4	4.6	–	–	–
$\sum_{i=1}^m \bar{e}_{ij}^v - e_i $	13.2	10.4	9.2	12.0	12.4	10.4	–	–	67.6

Source: Authors' own creation.

Table 4. The Results of the Expert Evaluation Concerning the Significance of the Indicators Representing the Phenomenon under Consideration on a 10-Point Scale

Indicator	Experts						$\sum_{j=1}^r e_{ij}$	$\bar{e}_i^h$	$\bar{e}_i^h - \sum_{j=1}^r e_{ij}$
	1	2	3	4	5	6			
1	10	2	8	3	6	4	33	5.5	15
2	9	3	7	1	6	2	28	4.67	16
3	8	4	9	3	7	4	35	5.83	13
4	7	2	10	4	8	3	34	5.67	16
5	8	3	9	2	9	2	33	5.5	19
$\sum_{j=1}^m r_j$	42	14	43	13	36	15	163	–	79
$\bar{r}_j^v$	8.4	2.8	8.6	2.6	7.2	3	–	–	–
$\sum_{j=1}^m e_j - \bar{e}_j $	4.4	3.2	4.4	4.4	5.2	4	8	–	25.6

Source: Authors' own creation.

The horizontal differentiation indicator of expert evaluations concerning the significance of the indicators representing the phenomenon under consideration is estimated as follows:

$$V^H = 1 - \frac{V_f^H}{V_{\max}^H},$$

here: V^H – horizontal differentiation indicator of expert evaluations; V_f^H – aggregate actual degree of horizontal differentiation; $V_{\max}^H$ – maximum possible degree of horizontal differentiation.

For estimation of value V_f^H , the average e_i^h of all expert evaluations of the i th indicator needs to be found:

$$\bar{e}_i^h = \frac{\sum_{j=1}^r e_{ij}}{r}.$$

It is now possible to estimate the magnitude of the total deviation from the average $\bar{e}_i^h$ of the expert evaluations concerning the significance of the i th indicator:

$$v_i^h = \sum_{j=1}^r |\bar{e}_i^h - e_{ij}|$$

here v_i^h – total deviation from the average $\bar{e}_i^h$ of the expert evaluations concerning the significance of the i th indicator.

The sum of values v_i^h will represent the total degree of horizontal differentiation of the expert evaluations (Tables 3 and 4):

$$V_f^H = \sum_{i=1}^m v_i^h$$

In our case, value V_f^H is equal to 16.7 and 79, respectively.

The essence of value $V_{\max}^H$ is that it indicates what maximum degree of horizontal differentiation of expert evaluations is possible for a particular number of indicators m and a particular number of experts r . With this in mind, the total deviation of all other system estimates from the value of a particular estimate $e_{ij}(i = \overline{1, 10})$ is found:

$$P_i = \sum_{j=1}^r |e_{ij} - e_{ij+1}|$$

here P_i – total deviation of all other system estimates from the value of the i th estimate.

For instance, if on a 10-point scale $e_i = 10$, then by subtracting all other system estimates we get that $P_{10} = 45$. Table 5 presents values P_i for all system estimates, i.e. when $m = 10$ and $r = 10$. The table is compiled considering the fact that in the $V_{\max}^H V_{\max}^H$ case, expert opinions are unanimous, i.e. $e_{ij} = e_{ij+1}$. Then, $e_{ij} = \bar{e}_i$.

Table 5. Deviations of the Sstem Estimates from each Individual Etimate

$\bar{e}(e_{ij})$	10	9	8	7	6	5	4	3	2	1	P_i	Rank
10		1	2	3	4	5	6	7	8	9	45	1–2
9	1		1	2	3	4	5	6	7	8	37	3–4
8	2	1		1	2	3	4	5	6	7	31	5–6
7	3	2	1		1	2	3	4	5	6	27	7–8
6	4	3	2	1		1	2	3	4	5	25	9–10
5	5	4	3	2	1		1	2	3	4	25	9–10
4	6	5	4	3	2	1		1	2	3	27	7–8
3	7	6	5	4	3	2	1		1	2	31	5–6
2	8	7	6	5	4	3	2	1		1	37	3–4
1	9	8	7	6	5	4	3	2	1		45	1–2

Source: compiled by the author.

In Table 5, the sum of the differences P_i for estimates 10 and 1 is equal, and thus the rank is the same (a similar situation can be observed when analysing other pairs of the estimates). Then, the question arises as to the order in which the estimates could be distributed among individual experts when their number is odd. For instance, which alternative is more acceptable when $r = 5$ – whether it is 1, 10, 1, 10, 1 or 10, 1, 10, 1, 10. For further estimations, the alternative with higher P_i must be accepted. The same is true for other indicators. Therefore, based on Table 5, a separate table for estimation of the size $V_{\max}$ can be constructed.

The table is constructed as follows. First, on the basis of the ranks provided in Table 5, the pairs of the values with most varying significance ranks are identified, for instance, 1–10, 2–9, 3–8, 4–7, 5–6. When filling the rows of the table, the estimates with the same rank are sorted. For example, if estimates 1–2 are ranked 1 and 10, and we have the total of 6 ranks, the first row of the table will be written as follows: 1, 10, 1, 10, 1, 10. The second row will be the opposite of the first: 10, 1, 10, 1, 10, 1. Other rows in the table are filled in the same way. The sum of the row ranks is then averaged, and the total deviation from the average is estimated (Table 6).

Table 6. The Table of the Expert Evaluations Concerning the Indicator Significance for Estimation of the size $V_{\max}$, when $m \leq 10$ and $r \leq 10$

Indicators	Experts										$\sum_{i=1}^r e_{ij}$	$\bar{e}$	P_i
	1	2	3	4	5	6	7	8	9	10			
1	1	10	1	10	1	10	1	10	1	10	55	5.5	45
2	10	1	10	1	10	1	10	1	10	1	55	5.5	35
3	2	9	2	9	2	9	2	9	2	9	55	5.5	25
4	9	2	9	2	9	2	9	2	9	2	55	5.5	15
5	3	8	3	8	2	8	3	8	3	8	55	5.5	5
6	8	3	8	3	8	3	8	3	8	3	55	5.5	5
7	4	7	4	7	4	7	4	7	4	7	55	5.5	15
8	7	4	7	4	7	4	7	4	7	4	55	5.5	25
9	5	6	5	6	5	6	5	6	5	6	55	5.5	35
10	6	5	6	5	6	5	6	5	6	5	55	5.5	45

Source: Authors' own creation.

When the number of experts is even, $V_{\max}$ is equal to:

$$V_{\max} = 4.5 \times m \times r.$$

In our case, i.e. when $m = 5.0$ and $r = 6$, $V_{\max}$ is equal to:

$$V_{\max} = 4.5 \times 5 \times 6 = 135.$$

The values of the size $S_{\max}^H$ for all cases, i.e. when $m = \overline{1, 10}$ and $r = \overline{1, 10}$, are presented in Table 7.

Table 7. The Values of the Size $S_{\max}^H$ ($m = \overline{1, 10}$, $r = \overline{1, 10}$)

Indicators	1	2	3	4	5	6	7	8	9	10
1	–	9	12	18	21,6	27	30.9	36	40	45
2	–	18	24	36	43	54	62	72	80	90
3	–	27	33	54	60	81	86	108	111	135
4	–	36	43	72	77	108	110	144	142	180
5	–	45	49	90	89	135	127	180	164	225
6	–	54	56	108	101	162	144	216	187	270
7	–	63	60	126	108	189	154	252	200	315
8	–	72	64	144	115	216	165	288	213	360
9	–	81	65	162	118	243	168	324	217	405
10	–	90	67	180	120	270	172	360	222	450

Source: Authors' own creation.

Based on Tables 3, 4 and 7, V^H for the examples under consideration is equal to 0.876 and 0.415, respectively.

The indicator of vertical differentiation of expert opinions is estimated as follows:

$$V^V = \frac{V_f^V}{V_{\max}}$$

here V^V – vertical differentiation indicator; V_f^V – aggregate actual degree of vertical differentiation; $V_{\max}$ – maximum possible degree of vertical differentiation.

For estimation of the size V_f^V , the average $\bar{e}_j^V$ of the estimations of all indicators provided by the i th expert needs to be found:

$$\bar{e}_j^V = \frac{\sum_{i=1}^m e_{ij}}{m}$$

It is now possible to estimate the total deviation of the j th expert's evaluations of the significance of all m indicators from average $\bar{e}_j^V$.

$$V_j^V = \sum_{i=1}^m |\bar{e}_j^V - e_i|$$

here V_j^V – total deviation of the j th expert's evaluations of the significance of all m indicators from average $\bar{e}_j^V$.

The sum of the sizes V_j^V will indicate the total degree of vertical differentiation of expert opinions:

$$V_f^V = \sum_{j=1}^r V_j^V$$

In our case, it is equal to 67.6 (Table 3) and 25.6 (Table 4), respectively.

The essence of the size $V_{\max}^H$ is that it indicates what is the maximum possible degree of vertical differentiation of expert opinions for a specific number of indicators m and a specific number of experts r . Thus, it is necessary to select such a set of indicators that would reflect the maximum differentiation of the vertical values. Based on Table 6, it can be expressed as follows (Table 8):

Table 8. Table of the Expert Evaluations for Estimation of the Size $V_{\max}^H$, when $m = 10$ and $r = 10$

Indicators	Experts									
	1	2	3	4	5	6	7	8	9	10
1	10	10	10	10	10	10	10	10	10	10
2	1	1	1	1	1	1	1	1	1	1
3	9	9	9	9	9	9	9	9	9	9
4	2	2	2	2	2	2	2	2	2	2
5	8	8	8	8	8	8	8	8	8	8
6	3	3	3	3	3	3	3	3	3	3
7	7	7	7	7	7	7	7	7	7	7

Indicators	Experts									
	1	2	3	4	5	6	7	8	9	10
8	4	4	4	4	4	4	4	4	4	4
9	6	6	6	6	6	6	6	6	6	6
10	5	5	5	5	5	5	5	5	5	5

Source: Authors' own creation.

Based on Formulas (7)–(9), the values of the size V_f^v for all alternatives in Table 8, i.e. when $m = \overline{1, 10}$ and $r = \overline{1, 10}$, can be estimated. The results of the estimations are provided in Table 9.

Table 9. Table of the Expert Evaluations for Estimation of the Size $V_{\max}^v$, when $m = 10$ and $r = 10$

Indicators	Experts									
	1	2	3	4	5	6	7	8	9	10
1	–	–	–	–	–	–	–	–	–	–
2	9	18	27	36	45	54	63	72	81	90
3	11	23	34	45	57	68	79	91	102	113
4	16	32	48	64	80	96	112	128	144	160
5	18	36	54	72	90	108	119	144	162	180
6	21	42	63	84	105	126	147	168	189	210
7	22	45	67	89	112	134	156	178	201	223
8	24	48	72	96	120	144	168	192	216	240
9	24	49	73	98	122	147	171	196	220	244
10	25	50	75	100	125	150	175	200	225	250

Source: Authors' own creation.

From Tables 7 and 9, it can be concluded that the degree of horizontal rather than vertical differentiation has a more significant impact on expert opinion consistency.

The values of the sizes V_f^H , V_f^V , $V_{\max}^H$ for all cases under consideration (Tables 1, 3 and 4) are presented in Table 10.

Table 10. The Results of the Expert Opinion Consistency Measuring

Example No.	The degree of expert opinion differentiation							Result
	horizontal			Vertical			total	
	V_f^H	$V_{\max}^H$	V^H	V_f^V	$V_{\max}^V$	V^V	V_f	
1 (Table 1)	59	135	0.56	0.37	108	0.37	0.21	Inconsistent opinions
2 (Table 3)	16.7	135	0.88	67.6	108	0.63	0.55	Consistent opinions
3 (Table 4)	79	135	0.42	25.6	108	0.24	0.10	Inconsistent opinions

Source: Authors' own creation.

In order to obtain a definitive answer as to the degree of expert opinion consistency in the cases under consideration, two following questions need to be answered. First, having which partial consistency indicators V^H and phenomena V^V , it can be stated that the degree of differentiation is sufficient in terms of expert opinion consistency; second, how the above-mentioned partial indicators can be combined into a single aggregate size.

To answer the first question, Kendall's expert opinion consistency measuring methodology will be employed. This methodology proposes that if the value of the indicator W is equal to or higher than 0.5, expert opinions are consistent.

Following this logic, the total degree of expert opinion consistency will be obtained by multiplying the values of both partial indicators:

$$V_f = V_f^H \cdot V_f^V$$

$$\tilde{V} = \tilde{V}^h \cdot \tilde{V}^v$$

here V – indicator of expert opinion consistency regarding significance of the indicators representing the phenomenon under consideration; $\tilde{V}^h$ – horizontal differentiation of significance of the indicators representing the phenomenon under consideration that satisfies the condition of consistency; $\tilde{V}^v$ – vertical differentiation of significance of the indicators representing the phenomenon under consideration that satisfies the condition of consistency.

Since $\tilde{V}^h = 0.5$ and $\tilde{V}^v = 0.5$, then

$$\tilde{V} = 0.5 \times 0.5 = 0.25$$

i.e. if the value of the indicator V of expert opinion consistency regarding significance of the indicators representing the phenomenon under consideration is equal to or higher than 0.25, expert opinions are considered consistent.

Of the three cases examined (Tables 1, 3 and 4), expert opinions are consistent only in the second case (Table 10).

4. Conclusions

When assessing the actual state of the development of particular SESs, not only the values, but also the weights of all indicators need to be considered. This is because SESs are highly complex and manifest in many different aspects. The indicators describing these aspects are ambiguous in their impact on a phenomenon under consideration, which is why their ultimate value is estimated as the product of their normalised value and significance. Because indicator significance is usually established through expert evaluation, the reliability of expert opinion consistency measurement is essential to the credibility of the resulting multi-criteria assessment.

The analysis confirms two main limitations of conventional rank-based consistency measurement. First, expert opinion consistency is assessed using

transformed ranks rather than the initial evaluations, which can reduce the information contained in the original scale. Second, the assessment focuses on the differentiation of aggregate indicator significance across experts, while differentiation within each expert's evaluations across indicators is not considered.

The methodology proposed in this article addresses these limitations by combining horizontal and vertical differentiation. Horizontal differentiation reflects variation in the evaluations of the same indicator across experts, whereas vertical differentiation captures variation across indicators within the evaluations of a single expert. Following the threshold logic used in Kendall's methodology, the partial indicators are considered satisfactory at values of 0.5 or higher, yielding a minimum aggregate consistency value of 0.25. In the three illustrative cases examined, only the second case satisfies this criterion, which is consistent with the pattern visible in the initial expert evaluations.

The practical examples therefore indicate that the proposed approach can provide a more informative assessment of expert opinion consistency when rank transformation masks differences in the original evaluations. At the same time, the method is illustrated here through a limited set of numerical examples; further validation using larger expert panels and empirical decision-making applications would help assess its robustness and generalisability.

Funding: The article was prepared as part of a project funded by the State Budget titled "Establishment of Centers of Excellence at Mykolas Romeris University," which is implemented under the initiative "Centers of Excellence Initiative" initiated by the Ministry of Education, Science and Sports of the Republic of Lithuania.

References

- [1] Balcerzak, A.P. (2016), *Multiple-criteria evaluation of quality of human capital in the European Union countries. Economics & Sociology*, 9(2), 11-26, <https://doi.org/10.14254/2071-789X.2016/9-2/1>.
- [2] Baležentis, T. (2022), *The kernel-based comprehensive aggregation PROMETHEE (PROMETHEE-KerCA) method for multi-criteria decision making with application to policy modelling. Journal of International Studies*, 15(1), 63-77, <https://doi.org/10.14254/2071-8330.2022/15-1/4>.
- [3] Behm, M. (2005), *Linking construction fatalities to the design for construction safety concept. Safety Science*, 43(8), 589-611, <https://doi.org/10.1016/j.ssci.2005.04.002>.
- [4] Brans, J.P., Mareschal, B., Vincke, P. (1984), *Promethee: A new family of outranking methods in multicriteria analysis*. In Brans, J.P. (Ed.), *Operational Research '84*. Elsevier Science Publishers B.V., North-Holland, Amsterdam, The Netherlands, 408-421.
- [5] Cheung, F.K.T., Kuen, L.L. (1991), *Multi-criteria evaluation model for the selection of architectural consultants. Construction Management and Economics*, 20(7), 569-581.
- [6] Churchman, C.W., Ackoff, R.L., Smith, N.M. (1954), *An approximate measure of value. Journal of the Operations Research Society of America*, 2(2), 172-187.

- [7] Dinca, V.M., Dima, A.M., Moagăr-Poladian, S., Săseanu, A.S., Dinu, V. (2025), *Energy consumption and savings behavior under the pressure of energy security, sustainability and cleaner energy transition: The case of Romania. Oeconomia Copernicana*, 16(2), 489-521, <https://doi.org/10.24136/oc.3749>.
- [8] Fang, D.P., Xie, F., Huang, X.Y., Li, H. (2004), *Factor analysis-based studies on construction workplace safety management in China. International Journal of Project Management*, 22(1), 43-49, [https://doi.org/10.1016/S0263-7863\(02\)00115-1](https://doi.org/10.1016/S0263-7863(02)00115-1).
- [9] Gedvilaitė, D. (2019), *The assessment of sustainable development of a country's regions. Doctoral dissertation*, Vilnius Gediminas Technical University. Technika, Vilnius, Lithuania.
- [10] Ginevicius, R., Nazarko, J., Gedvilaite, D., Petraskevicius, V. (2023), *Ranking the importance of multi-criteria evaluation indicators based on their transitivity. Economics & Sociology*, 16(4), 309-320, <https://doi.org/10.14254/2071-789X.2023/16-4/15>.
- [11] Ginevičius, R., Trišč, R., Remeikienė, R., Zielińska, A., Strikaitė-Latušinskaja, G. (2022), *Evaluation of the condition of social processes based on qualimetric methods: The COVID-19 case. Journal of International Studies*, 15(1), 230-249, <https://doi.org/10.14254/2071-8330.2022/15-1/15>.
- [12] Hwang, C.L., Lin, M.J. (1987), *Group Decision Making under Multiple Criteria: Methods and Applications*. Springer, Berlin-Heidelberg, Germany.
- [13] Hwang, C.-L., Yoon, K. (1981), *Multiple Attribute Decision Making: Methods and Applications: A State-of-the-Art Survey*. Springer-Verlag, New York, USA.
- [14] Ivanová, E., Masárová, J. (2016), *Assessment of innovation performance of Slovak regions. Journal of International Studies*, 9(2), 207-218, <https://doi.org/10.14254/2071-8330.2016/9-2/16>.
- [15] Kartam, N.A., Flood, I., Koushki, P. (2000), *Construction safety in Kuwait: Issues, procedures, problems, and recommendations. Safety Science*, 36(3), 163-184, [https://doi.org/10.1016/S0925-7535\(00\)00041-2](https://doi.org/10.1016/S0925-7535(00)00041-2).
- [16] Kendall, M.G. (1948), *Rank Correlation Methods*. Griffin, London, UK.
- [17] Kendall, M.G. (1975), *Rank Correlation Methods*, 4th ed. Charles Griffin, London, UK.
- [18] Keshavarz, G.M., Zavadskas, E.K., Olfat, L., Turskis, Z. (2015), *Multi-Criteria Inventory Classification Using a New Method of Evaluation Based on Distance from Average Solution (EDAS). Informatica*, 26(3), 435-451.
- [19] Khodaparasti, R.B., Aboulfazli, A., Isakhajelou, R. (2015), *Ranking the most effective marketing mix elements on the sales of Javid Darb company products: An AHP technique. Journal of International Studies*, 8(2), 164-173, <https://doi.org/10.14254/2071-8330.2015/8-2/14>.
- [20] Klee, A.J. (1971), *The role of decision models in the evaluation of competing environmental health alternatives. Management Science*, 18(2), B52-B67, <https://www.jstor.org/stable/2629528>.
- [21] Kostiukevych, R., Melnyk, L., Krulický, T., Kostiukevych, A., Melnyk, L. (2020), *A value-oriented approach to assessing regional economic development projects. Journal of International Studies*, 13(4), 155-170, <https://doi.org/10.14254/2071-8330.2020/13-4/11>.

- [22] Libby, R., Blashfield, R.K. (1978), *Performance of composite as a function of the number of judges*. *Organizational Behavior and Human Performance*, 21(2), 121-129.
- [23] Łopatka, P., Thuczak, A., Jadach, R. (2023), *Multi-criteria decision support methods for the selection of the optimum contractor: A case study of a passenger transfer centre*. *Journal of Sustainable Development of Transport and Logistics*, 8(2), 219-228, <https://doi.org/10.14254/jsdtl.2023.8-2.16>.
- [24] MacCrimmon, K.R. (1968), *Decisionmaking Among Multiple-Attribute Alternatives. A Survey and Consolidated Approach*. RAND Memorandum RM-4823-ARPA, https://www.rand.org/pubs/research_memoranda/RM4823.html.
- [25] Mateiaș, D.E., Kanungo, S. (2025), *Patterns of Sustainable Decision-Making Based on Decision Biases and Professional Group Performance*. In: Dima, A.M., Badarinza, C. (eds), *Innovative Approaches in Economics, Leadership, and Technology. ICESS 2024. Springer Proceedings in Business and Economics*. Springer, Cham, Switzerland, https://doi.org/10.1007/978-3-031-86989-1_10.
- [26] Mubarak, E.M., Zuhair Ridha, A., Abdullah, Z.T. (2024), *Multi-criteria assessment methodology for remanufacturing conventional grinding machines into CNC machine tools*. *Economics. Management and Sustainability*, 9(2), 29-43, <https://doi.org/10.14254/jems.2024.9-2.3>.
- [27] Opricovic, S., Tzeng, G.-H. (2004), *Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS*. *European Journal of Operational Research*, 156(2), 445-455, [https://doi.org/10.1016/S0377-2217\(03\)00020-1](https://doi.org/10.1016/S0377-2217(03)00020-1).
- [28] Oželienė, D. (2019), *Modelling the factors of a company's sustainable development*. Doctoral dissertation, Vilnius Gediminas Technical University. Technika, Vilnius, Lithuania.
- [29] Podvezko, V. (2005), *Agreement of expert estimates*. *Technological and Economical Development of Economy*, 11(2), 101-107.
- [30] Roy, B. (1968), *Revisiting carrying capacity: Area-based indicators of sustainability (la methode Electre)*. *Revue Francaise d'informatique et de Recherche Operationnelle*, 8, 57-75.
- [31] Saaty, T.L. (2001), *Fundamentals of the analytic hierarchy process*. In Schmoldt, D.L., Kangas, J., Mendoza, G.A., Pesonen, M. (eds), *The Analytic Hierarchy Process in Natural Resource and Environmental Decision Making*, 15-35, Springer, Dordrecht, Netherlands.
- [32] Šaparauskas, J. (2004), *Darnaus miesto vystymo(-si) daugiatiakslė selektonovacija [Multi-objective selective innovation for sustainable urban development]*. Doctoral dissertation, Vilnius Gediminas Technical University. Technika, Vilnius, Lithuania.
- [33] Srinivasan, V., Shocker, A.D. (1973), *Estimating the weights for multiple attributes in a composite criterion using pairwise judgments*. *Psychometrika*, 38, 473-493, <https://doi.org/10.1007/BF02291490>.
- [34] Stankevičienė, J., Maditinos, D.I., Kraujalienė, L. (2019), *MULTIMOORA as the instrument to evaluate the technology transfer process in higher education institutions*. *Economics & Sociology*, 12(2), 345-360, <https://doi.org/10.14254/2071-789X.2019/12-2/21>.

- [35] Štreimikienė, D. (2023), *Waste management in Baltic States: Comparative assessment. Journal of International Studies*, 16(4), 39-51, <https://doi.org/10.14254/2071-8330.2023/16-4/3>.
- [36] Triantaphyllou, E. (2000), *Multi-Criteria Decision Making Methods*. Kluwer Academic Publishers, Bonn, Germany, <https://doi.org/10.1007/978-1-4757-3157-6>.
- [37] Ubartė, I. (2017), *Daugiakriterė sprendimų paramos ir rekomendacijų sistema sveikam ir saugiam būstui užstatytoje aplinkoje vertinti [Multiple Criteria Decision Support and Recommender System for the Assessment of Healthy and Safe Homes in the Built Environment]. Doctoral dissertation*, Vilnius Gediminas Technical University. Technika, Vilnius, Lithuania.
- [38] Ustinovičius, L. (2003), *Statybos investicijų efektyvumo nustatymo sprendimų paramos sistema [Construction investment efficiency measuring solution support system]. Habilitation doctoral thesis*. Technika, Vilnius, Lithuania.
- [39] Voda, A.I., Șoitu, D.T., Cojocaru, D.C. (2025), *Exploring the Dynamics of Knowledge Ecosystem: A Systematic Literature Review*. In: Dima, A.M., Badarinza, C. (eds), *Innovative Approaches in Economics, Leadership, and Technology. ICES 2024. Springer Proceedings in Business and Economics*. Springer, Cham, Switzerland, https://doi.org/10.1007/978-3-031-86989-1_2.
- [40] Volkov, A. (2018), *Assessment of the impact of the common agricultural policy direct payments system on agricultural sustainability. Doctoral dissertation*, Vilnius Gediminas Technical University. Technika, Vilnius, Lithuania.
- [41] Wątróbski, J., Jankowski, J., Ziemba, P. (2016), *Multistage performance modelling in digital marketing management. Economics & Sociology*, 9(2), 101-125, <https://doi.org/10.14254/2071-789X.2016/9-2/7>.
- [42] Yoon, K. (1980), *Systems selection by multiple attribute decision making. Ph.D. dissertation*. Kansas State University, Kansas, USA, <https://trove.nla.gov.au/version/20403381>.
- [43] Zagorskas, J., Burinskienė, M., Zavadskas, E.K., Turskis, Z. (2007), *Urbanistic assessment of city compactness on the basis of GIS applying the COPRAS method. Ekologija*, 53, 55-63.
- [44] Zavadskas, E.K., Kaklauskas, A. (1996), *Pastatų sistemos techninis vertinimas [Building systemic-technical assessment]*. Technika, Vilnius, Lithuania.
- [45] Zavadskas, E.K., Turskis, Z., Tamošaitienė, J., Marina, V. (2008), *Multicriteria selection of project managers by applying grey criteria. Technological and Economic Development of Economy: Baltic Journal on Sustainability*, 14(4), 462-477, Technika, Vilnius, Lithuania.