Manuscript of / please cite as
Bozóki, S., Dezső, L., Poesz, A., Temesi, J. [2013]:
Analysis of pairwise comparison matrices: an empirical research,
Annals of Operations Research 211(1), pp.511-528.
DOI 10.1007/s10479-013-1328-1 http://dx.doi.org/10.1007/s10479-013-1328-1

Analysis of pairwise comparison matrices: an empirical research[*]

Sándor Bozóki

Computer and Automation Research Institute, Hungarian Academy of Sciences (MTA SZTAKI); Department of Operations Research and Actuarial Sciences, Corvinus University of Budapest, Hungary

Linda Dezső, prae-doc
Department of Applied Psychology: Work, Education and Economy, University of Vienna, Austria

Attila Poesz, PhD Student

Department of Operations Research and Actuarial Sciences, Corvinus University of Budapest, Hungary

József Temesi

Department of Operations Research and Actuarial Sciences, Corvinus University of Budapest, Hungary

e-mail: jozsef.temesi@uni-corvinus.hu

Abstract: Pairwise comparison (PC) matrices are used in multi-attribute decision problems (MADM) in order to express the preferences of the decision maker. Our research focused on testing various characteristics of PC matrices. In a controlled experiment with university students (N = 227) we have obtained 454 PC matrices. The cases have been divided into 18 subgroups according to the key factors to be analyzed. Our team conducted experiments with matrices of different size given from different types of MADM problems. Additionally, the matrix elements have been obtained by different questioning procedures differing in the order of the questions. Results are organized to answer five research questions. Three of them are directly connected to the inconsistency of a PC matrix. Various types of inconsistency indices have been applied. We have found that the type of the problem and the size of the matrix had impact on the inconsistency of the PC matrix. However, we have not found any impact of the questioning order. Incomplete PC matrices played an important role in our research. The decision makers behavioral consistency was as well analyzed in case of incomplete matrices using indicators measuring the deviation from the final order of alternatives and from the final score vector.

Keywords: multi-attribute decision making; experimental techniques in decision making; pairwise comparisons; inconsistency; incomplete pairwise comparison matrix

1 Introduction

In additive models of multi-attribute decision making it is assumed that alternatives can be evaluated criterion-wise, and their weighted sum – where weights represent the importance of the criteria –, yields an appropriate numerical

[*] The research was supported by the Hungarian Scientific Research Fund, Grant K 77420.

assessment of the overall performance of the alternatives. Total scores of the alternatives help the decision maker to choose the best one, or to rank them. That approach includes two tasks: to compute attributes' weights and evaluate each alternative with respect to each criterion. First, we need to express the importance of attributes by numbers based on decision maker's preferences (weights). Second, alternatives need to be evaluated numerically with respect to each criterion (preferences, scores). Pairwise comparison matrices are applied in both tasks. In this article we will use the terminology of the second task, that is, having a criterion fixed, the aim is to evaluate the alternatives.

Let $v_1, v_2, \ldots, v_n > 0$ denote the scores of the alternatives according to a given criterion. v_i-s are also called preference values and their sum is usually normalized to 1. Value of v_i ($i = 1,2,\ldots,n$) is usually not known explicitly, but we assume that the decision maker is able to approximate the ratios v_i/v_j by answering a question like 'How many times alternative i is better than (preferred to) alternative j according to the given criterion?' Having the answers for all pairs an $n \times n$ matrix can be built. Matrix $\mathbf{T} = [t_{ij}]$ contains the revealed ratios and it is called pairwise comparison (PC) matrix and t_{ij} is the approximation of v_i/v_j for all i,j. $\mathbf{T}$ is a positive reciprocal matrix, where the diagonal elements are equal to 1 ($t_{ij} > 0$ and $t_{ij} = 1/t_{ji}$, for all $i, j = 1,\ldots, n$). The reciprocal property can be ensured if the decision maker gives only one of the ratios from the pairs (t_{ij}, t_{ji}) and the other one will be calculated using the reciprocity rule.

Considering a perfect (or informed) decision maker, who can precisely estimate the unknown ratios – which case is very exceptional – $\mathbf{T}$ is a consistent matrix: $t_{ik} * t_{kj} = t_{ij}$, for all $i, j, k = 1,\ldots,n$ (chain rule). For most real problems, however, $\mathbf{T}$ is not consistent and its inconsistency is accounted for the behavior of the decision maker *(Choo and Wedley, 2004; Temesi, 2011)*. A consistent matrix has rank one, and it is generated by any of its columns *(Saaty, 1980)*. In this case the normalized (sum equals to 1) values of v_i ($i = 1,\ldots,n$), can be calculated from any column of matrix $\mathbf{T}$ by $v_i = t_{ij}$ / sum(t_{kj}, $k = 1,2,..,n$), where $1 \leq j \leq n$ is an arbitrary integer.

If $\mathbf{T}$ is not consistent and values t_{ij} are considered as the approximation of the ratios v_i/v_j, then several estimation methods can solve the problem. If the ratios are given applying a 1,..,9 ratio scale and the eigenvalue method is used for the estimation of the values v_i, we arrive at a special case: Saaty's Analytic Hierarchy Process – AHP *(Saaty, 2005)*. This paper neither deals with the characteristics of the estimation methods, nor will be AHP in the forefront of the discussion. Instead, our research focuses on the properties of the PC matrix in the process of generating the elements of the PC matrix.

Elicitation of the elements of the PC matrix is implemented by a questioning procedure, in which the decision maker answers the pairwise comparison questions one by one. If matrix $\mathbf{T}$ is not consistent then the biases in the comparisons can be classified into different types reflecting on the potential sources of the deviation from the consistent case. Real-world applications rarely describe how the elements of the PC matrix have been generated, and do not analyze the reasons of inconsistency, and the choice of a particular consistency improving method. One of the aims of our research was to explore the characteristics of empirical PC matrices and their relationship to inconsistency.

Incomplete pairwise comparison matrices are also in the focus of our interest. The estimation method could use an incomplete PC matrix, for instance, if the questioning process was interrupted or elements are missing from a complete PC matrix without a possibility to replace them. Several studies dealt with incomplete PC matrices from different aspects *(Harker, 1987; Bozóki, Fülöp and Rónyai, 2010)*. Optimal completion of these matrices and inconsistency calculations using incomplete PC matrices are especially relevant for our discussion. An important research question is if incomplete PC matrices could be used for reducing the number of pairwise comparisons.

This paper presents results of an experiment on PC matrices conducted to investigate the characteristics of empirical PC matrices. Although previous research has shown the need for a large number of well-documented matrices obtained from a controlled environment, only a few studies have been analyzed empirical PC matrices. *Gass and Standard (2002),* for instance, highlighted important differences between randomly generated and experimental matrices. *Poesz (2008)* collected PC matrices from reported real-world applications to analyze their characteristics and drew conclusions about their inconsistency levels. *Linares (2009)* investigated the inconsistency of experimental pairwise comparison matrices.

The main goal of our research was to analyze the properties of PC matrices with the help of a database which was derived from our experiments. Prior to designing and running the experiments five research questions were formulated. At first, we describe the experiment, then present the results. This will be followed by drawing conclusions and proposing directions for further research.

2 The experiment

The experiment was conducted in 2010 at Corvinus University of Budapest, Hungary. 227 undergraduate and graduate students participated in the experiments. Subjects were 3rd year Bachelor and 1st year Master students of business and economics majors. The mean age was 22, where a low standard deviation reflected having students as subjects. 39% of the subjects were males, and 61% were females. This skewed gender distribution is consistent with the gender distribution of the students at Corvinus University.

The experiments were conducted in class as previously arranged with the professors. One session lasted approximately 25 minutes. First, the professor introduced the experimenters to the students, and he announced that participation is voluntary and could be discontinued at any time. Note, that no student refused participation in any of the sessions.

Participants received the experimental material in a stapled leaflet with a unique identification number. Each page of the leaflet displayed one comparison, thus each comparison was displayed on a separate page; subjects were not allowed to turn back pages. The first page was a practice task, after which students were encouraged to ask (if any) questions. When they finished working they were asked to wait until everyone was done. Each session was closed with debriefing.

In each session the experiment consisted of two subsequent tasks that were designed to test our hypotheses. The design of the test problems and the experimental setting captured the following four dimensions for future analysis:

- type of the problem,
- size of the PC matrix,
- questioning order,
- completeness.

In order to investigate the impact of the *type (nature) of the problem* the quality of the applied stimuli was categorized into "subjective" and "objective" groups. For the objective stimuli, our subjects were asked to compare countries by their size. First, subjects had to indicate from a presented pair of countries which country is larger. Then, they had to indicate by how much it is larger on a numerical scale. Thus, if one country was judged 30% larger than the other one, it was indicated to be 1.3 times larger. For the subjective stimuli subjects had to compare summer houses. At first they were asked to reveal which summer house they liked more. Then they indicated how much more they liked the preferred house on a numerical scale. For this latter they were as well given a Saaty-scale *(Saaty, 1980).* Thus, each comparison consisted of a dichotomous, verbal comparison and a subsequent estimation on a numerical scale. Note, that *Bozóki and Rapcsák (2008)* showed that using Saaty's inconsistency index it is irrelevant if either discrete or continuous scales was used.

Every subject was presented with one subjective (summer houses task) and one objective (country maps task) set of stimuli and the order of the presentation (that is whether subjective stimuli was given first or second) was randomly assigned to each subject. The countries and the summer houses were projected on the screen, which is a usual classroom practice. Note, that we used an imaginary map with irregular contours of the countries.

The second factor administered in the experiments was the *size of the matrices*. We applied three sets of matrices with the size 4×4, 6×6 and 8×8. Note, that in one session only one size was applied.

When the elements of the PC matrix are determined one by one, we refer to this as a questioning procedure. The third factor was the impact of three different questioning procedures differing in *the order of the questions*. In the first procedure subjects compared the summer houses and the countries in a sequential order. Country A, for instance, was first compared to country B, then country A was compared to country C, etc. In the second procedure the subjects compared the summer houses and the countries in a random order. For the third procedure we applied the order proposed by Ross *(Ross, 1934).* This procedure satisfies two conditions of an optimally balanced comparison. On one hand, it maximizes the distances for the same items to reappear. On the other hand, for every item the number of the first and the second positions in the comparison should be the closest possible. In contrast to sequential order, where, e.g., country A appears in each of the first five questions and it is always in the first position, Ross order balances both the frequency of reappearing and the first/second positions as much as possible. *Table 1* presents examples of filling out 6×6 matrices in each of the detailed questioning orders.

Table 1 A completed 6×6 matrix applying the three questioning orders

question / order	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
sequential	A-B	A-C	A-D	A-E	A-F	B-C	B-D	B-E	B-F	C-D	C-E	C-F	D-E	D-F	E-F
random	A-F	B-E	A-C	F-E	C-D	B-D	B-F	A-E	C-E	A-D	E-D	C-F	B-C	A-D	B-A
Ross	A-B	F-D	E-A	C-B	E-F	A-C	B-D	F-A	D-C	E-B	A-D	C-E	B-F	D-E	C-F

The experiment was a 2(types)×3(sizes)×3(questioning orders) factorial design determined by the three factors described above. There were 9 sessions run all together, with 25 participants on average in each session. Every subject received a set of objective and subjective type of stimuli. Thus, we ended up having a total number of 454 complete PC matrices (see *Table 2* for details).

Table 2 The number of subjects participated in separate experiments

experiment type / number of alternatives	objective	subjective	total
	230	224	454
4	68	69	137
6	80	77	157
8	82	78	160
questioning order			
sequential	75	75	150
random	77	74	151
Ross	78	75	153

When we checked the data we have found that some answer sheets contained either missing or obviously incorrect comparisons, therefore the final number of comparisons in our analysis were less, altogether from 445 datasheets.

Furthermore, the experimental design allowed us to trace and to analyze the properties of *incomplete PC matrices*. We have recorded every stage of completing the PC matrices in the course of a given questioning order. This way we have obtained a pool of incomplete matrices, which allowed us to investigate the characteristics of the incomplete matrices.

3 Results

3.1 Consistency analysis

The low level of inconsistency of a PC matrix is a necessary condition to obtain the right results when scores, weights, or preferences are obtained from the PC matrix. Using the definition of the consistent PC matrix several inconsistency indices can be developed.

The well-known Saaty index (CR – consistency ratio, *Saaty, 1980*) is based on the fact that the dominant eigenvalue of a consistent PC matrix is n. In general, CR is a positive linear transformation of the Perron eigenvalue λ_{max} as follows: CR = $(\lambda_{max} - n)/(RI_n*(n-1))$, where RI_n is defined as $(\Lambda_{max} - n)/(n-1)$, where Λ_{max} is an average value of the Perron eigenvalues of randomly generated $n \times n$ PC matrices. CR is zero if and only if the PC matrix is consistent, otherwise CR is positive. CR is widely used and a threshold value of 0.1 (10%) has been accepted in the practice. However, the concept of the CR is being heavily debated *(Murphy, 1993; Bana e Costa and Vansnick, 2008)*. One of the drawbacks of the index arises from its construction: having RI_n in the formula CR could not be investigated and interpreted analytically. It has been computationally verified that RI_n depends not only on n, but as well as on the maximal value of the ratio scale, however, it is ultimately irrelevant whether the ratio scale is discrete or continuous *(Bozóki and Rapcsák, 2008)*.

Yet another inconsistency index, based on the properties of 3×3 consistent PC matrices was developed by *Koczkodaj (1993)*. The 3×3 size PC matrix, called triad, can be written as follows:

$$\begin{pmatrix} 1 & a & b \\ 1/a & 1 & c \\ 1/b & 1/c & 1 \end{pmatrix}$$

CM inconsistency of a triad is based on the chain rule and defined as

$$CM(a,b,c) = \min\left\{ \frac{1}{a}\left| a - \frac{b}{c} \right|, \frac{1}{b}\left| b - ac \right|, \frac{1}{c}\left| c - \frac{b}{a} \right| \right\}$$

Furthermore, the extension of CM for higher dimensions *(Bozóki and Rapcsák, 2008)* can be written for an $n \times n$ PC matrix A by using the maximum function over the set of the triads as

$$\mathrm{CM}(\mathrm{A}) = \max\left\{ \mathrm{CM}(a_{ij}, a_{ik}, a_{jk}) \;\middle|\; 1 \le i < j < k \le n \right\}.$$

One of the advantages of CM is that it localizes the 'worst' triad in the PC matrix, an appropriate reconsideration of which (if it is possible) decreases the level of inconsistency of the whole PC matrix. It has also been shown that CR and CM are equivalent for 3×3 PC matrices due to a function-like relation of the two indices *(Bozóki and Rapcsák, 2008)*. However, for larger sizes this relation is not one-to-one. CM always ranges from 0 (in case of consistency) to 1; however, intermediate values are not translated into categories. One of the particular disadvantages of CM is that – up to this point – the threshold for acceptance has yet neither been defined nor validated.

Several other inconsistency indices have also been developed (find a detailed list of inconsistency indices by *Brunelli and Fedrizzi (2011)*. *Gass and Rapcsák (2004)* developed for instance an inconsistency measure, which is based on singular value decomposition. To represent the two different approaches we used the Saaty index (CR) and the Koczkodaj index (CM) in our analysis.

The first research question focused on the magnitude of inconsistency regarding different types of the decision problems: *Does subjective stimuli yields higher inconsistency level than objective stimuli?* We predicted that CR indices will be higher for subjective stimuli (i.e., summer houses task) than for objective stimuli (i.e., maps task).

We computed CR and CM inconsistency indices for all obtained complete PC matrices. *Table 3* presents the CR averages and *Table 4* shows the CM averages for all of the matrices, where the discriminating factor was the type of the problem. In both tables a single cell represents the average of 22-27 matrices.

As it can be seen from *Table 3* average CR indices are close to 10% for summer houses task (with the highest value of 13.31% for the 8×8 matrices) while they range around 1% for the maps tasks.

Table 3 The average of CR indices (in %) for complete matrices

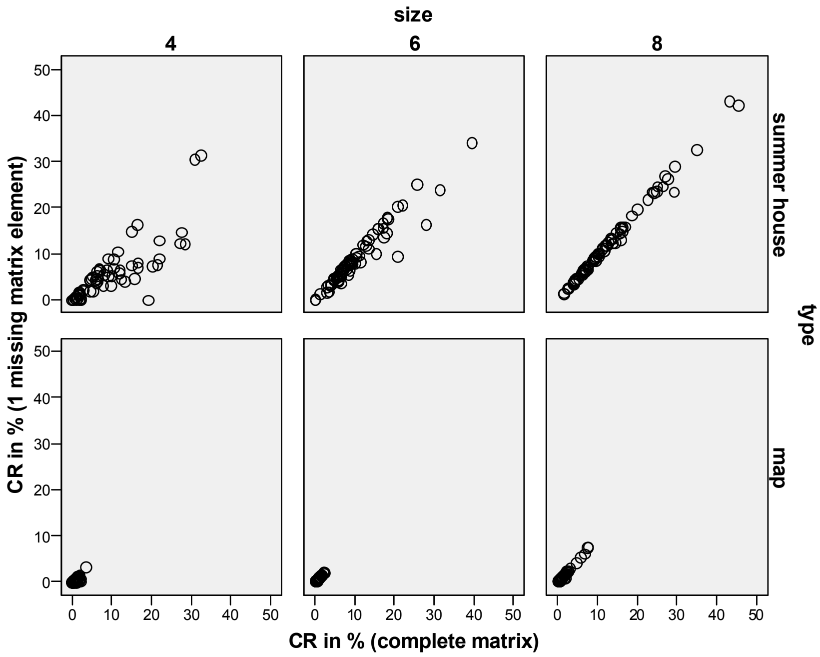
	CR (in %)					
size / order	summer houses			maps		
	4x4	6x6	8x8	4x4	6x6	8x8
sequential	8.10	10.75	12.46	0.67	0.81	1.28
random	10.38	9.47	11.96	0.78	0.80	1.07
Ross	8.75	10.63	13.31	0.70	0.88	1.73
total	9.06	10.28	12.58	0.72	0.83	1.36

As we predicted we found that CR indices (disregarding order and size) are higher for summer houses than for maps, Mann-Whitney U Test = 2285.00, $p \leq .001$.

From the same table one can see that for the summer houses the decision makers were close to Saaty's 10% acceptance rule for each size and for each questioning order. This high level of inconsistency was also found for the individual cases, as it can be seen in *Figure 1.*

Figure 1 visualizes the individual CR values for complete 4×4, 6×6 and 8×8 matrices and the CR values for the same matrices with one missing element. This figure also exemplifies the dynamics of the CR values of incomplete PC matrices using elementary data. Projecting the two-dimensional data to one of the axis highlights the distribution of the individual CR values.

Figure1 Individual CR values (in %)



This figure denotes the individual CR values with circles for summer houses (top row) and for maps (bottom row) for matrices with different sizes (columns). Both axes present the CR values from 0% to 40%: the horizontal axis for the complete matrices, the vertical axis for those incomplete matrices with one missing element (the last one in the questioning order). At this point we focus on mapping of the individual indices to the horizontal axis, i.e. on CR values of the complete matrices. (Note, that later in the paper this figure will help us demonstrating some properties of the incomplete matrices.) The figure shows that the CR values of the objective type are concentrated in the proximity of 0.

Furthermore, *Table 4* presents the average CM indices for the same matrices. The CM values are consistent with the tendencies obtained from *Table 3*. Interpretation of the magnitude of the CM values and their order can be interpreted similarly to the CR indices, but we have to note that CM threshold for acceptance has only been defined for a very special case *(Koczkodaj, Herman and Orlowski, 1997).*

Table 4 The average of CM indices for complete matrices

	CM					
size	summer houses			maps		
order	4x4	6x6	8x8	4x4	6x6	8x8
sequential	0.62	0.79	0.87	0.29	0.45	0.54
random	0.68	0.77	0.86	0.31	0.45	0.57
Ross	0.60	0.82	0.90	0.28	0.45	0.58
total	0.63	0.79	0.88	0.29	0.45	0.56

In the literature on MADM methodology most scholars and practitioners agree there is a decrease in the decision maker's consistency as the size of the matrix increases. Some researchers follow the rule of "magical number seven" *(Miller, 1956; Saaty 2003)* suggesting that matrices larger than 7×7 inherently lead to inconsistency, since the great number of questions imposes an overwhelming cognitive load on the decision maker: an 8×8 size matrix, for instance, includes 28 comparisons. One might argue that focusing on a single pair at a time should not necessarily lead to increased inconsistency; but in the end, increasing the matrix size by just one substantially increases the number of comparisons, making the entire process longer and more tiresome, possibly leading to diminished performance. These considerations led us to another research question: *Does increased size predict higher inconsistency level?*

In *Table 3* we earlier showed that an increase in size is associated with an increase in CR index, and *Table 4* also showed similar pattern for CM index. Regressing the logarithm of CR index on type, order, size and the type×size interaction, we see that the summer house task has a higher mean CR index than the map task, and that increasing matrix sizes lead to the same linear increase in CR index for both tasks (see parameter estimates for this regression in *Table 5*). Results of the regression of the logarithm of CM index on the same predictors are similar, with the exception that the type×size interaction is significant. So although both tasks exhibit increasing CM indices with larger matrices, the increase in the summer house task (which again has a higher intercept than the map task) is less severe (See *Table 6* for parameter estimates).

Table 5 Parameter estimates for linear regression of log_CR index

	Estimate	Std error	Wald chi-sq (df)	p-value
type				
map	0	N/A	N/A	.N/A
summer house	2.34	.32	53.13 (1)	.00
order				
random	0	N/A	N/A	N/A
sequential	-.02	.10	.05 (1)	.83
Ross	.04	.10	.15 (1)	.70
size	.15	.04	16.51 (1)	.00
summer house×size	.00	.05	.00 (1)	.95

Overall likelihood ratio chi-sq(5) = 477.01, p-value < 0.001

Table 6 Parameter estimates for linear regression of log_CM index

	Estimate	Std error	Wald chi-sq (df)	p-value
type				
map	0	N/A	N/A	N/A
summer house	1.06	.11	85.09 (1)	.00
order				
random	0	N/A	N/A	N/A
sequential	-.01	.04	.03 (1)	.86
Ross	.01	.04	.04 (1)	.84
size	.16	.01	162.63 (1)	.00
summer house×size	-.07	.02	16.51 (1)	.00

Overall likelihood ratio chi-sq(5) = 392.66, p-value < 0.001

Based on the analyses we have two propositions answering our two research questions.

Proposition 1: The level of inconsistency for the subjective tasks will be systematically greater than for the objective tasks.

Proposition 2: Inconsistency increases as the size of the PC matrix increases.

All in all, our findings reinforce what has already been identified and reported in the literature that inconsistency increases as the size of the matrix increases. Additionally, we found that increasing size results in greater inconsistency for subjective type of task than for objective type of task.

Although one of the novelties of our research was to test whether there are any differences in the consistency elicited by the different questioning orders, we have not investigated yet the impact of the questioning order: *Does the questioning order influence inconsistency?*

We conjectured that one of the questioning orders (e.g. the sequential method) might lead to lower inconsistency than the others. However, data presented in *Table 3* and in *Table 4* do not support our intuition: *Table 5* and *Table 6* further confirm that the questioning order does not have predictive power.

Proposition 3: The questioning order does not influence inconsistency.

3.2 Analyzing the incomplete matrices

In the further analysis we will use the generalization of PC matrix to the incomplete case. Incomplete matrices give an insight to the process of eliciting the elements of the complete PC matrix.

According to *Harker (1987)* an incomplete PC matrix contains one or more missing elements. Note that every (complete) PC matrix is built up through the series of incomplete PC matrices in the following fashion: whenever the decision maker enters a matrix element there will be one less missing element and this continues until the PC matrix has no missing values, i.e., becomes complete. CM inconsistency index, the Eigenvector Method and the CR inconsistency index have been extended to incomplete PC matrices *(Bozóki, Fülöp and Koczkodaj, 2011; Bozóki, Fülöp and Rónyai, 2010)* respectively. We used algorithms introduced by these two papers.

Let **A** be an incomplete pairwise comparison matrix where the missing (i.e., unknown) elements are denoted by *.

$$\mathbf{A} = \begin{array}{|c|c|c|c|c|} \hline 1 & a_{12} & * & .. & a_{1n} \\ \hline 1/a_{12} & 1 & a_{23} & .. & * \\ \hline * & 1/a_{23} & 1 & .. & a_{3n} \\ \hline : & : & : & & : \\ \hline 1/a_{1n} & * & 1/a_{3n} & .. & 1 \\ \hline \end{array}$$

Let $x_1, x_2 ,\ldots, x_d > 0$ denote the missing elements in the upper triangular part of **A**, and their reciprocals, $1/x_1$, $1/x_2$, …, $1/x_d$ are written in the lower triangular part of **A**. Known elements are denoted by a_{ij} as usual, except for the positions where entries are missing. $\mathbf{A}(x_1, x_2 ,\ldots, x_d)$ is a complete pairwise comparison matrix for any values of $(x_1, x_2 ,\ldots, x_d)$. The total number of missing elements in matrix **A** is *2d*.

$$\mathbf{A}(x_1, x_2 ,\ldots, x_d) = \begin{array}{|c|c|c|c|c|} \hline 1 & a_{12} & x_1 & .. & a_{1n} \\ \hline 1/a_{12} & 1 & a_{23} & .. & x_d \\ \hline 1/x_1 & 1/a_{23} & 1 & .. & a_{3n} \\ \hline : & : & : & & : \\ \hline 1/a_{1n} & 1/x_d & 1/a_{3n} & .. & 1 \\ \hline \end{array}$$

Based on the correspondence between the *CR* inconsistency and λ_{max} of a pairwise comparison matrix, the generalization of the Eigenvector Method for the incomplete case is originated from the optimal solution of the eigenvalue minimization problem as follows:

$$\lambda^*_{max} = \min \{ \lambda_{max} (\mathbf{A}(x_1, x_2 ,\ldots, x_d)) \mid x_1, x_2 ,\ldots, x_d > 0 \}.$$

Now CR can be extended for the incomplete case by using the same definition as in section 3.1:

$$\mathrm{CR}(\mathbf{A}) = (\lambda^*_{max} - \mathrm{n})/(\mathrm{RI}_n^*(n - 1)).$$

Analogously, the extension of CM for the incomplete case is

$$\mathrm{CM}(\mathbf{A}) = \min \{ \mathrm{CM}(\mathbf{A}(x_1, x_2 ,\ldots, x_d)) \mid x_1, x_2 ,\ldots, x_d > 0 \}.$$

Since we recorded every entry that the decision makers made, it is possible to track and analyze the change (if any) in the behavioral consistency of the decision maker. Thus, we can measure/index the inconsistency throughout the procedure and locate the inconsistency. *Table 7* and *Table 8* present results of such an analysis. The "number of matrix elements" represents the (ordinal) answer number in the sequence. For a 6×6 matrix this range is from 5 to $n(n - 1)/2 = 15$. The average CR inconsistencies were computed for each stage in the sequence broken down by questioning order. Our research question now: *Is the behavior of the decision maker consistent in the course of the entire questioning procedure?*

Table 7 The average of CR inconsistencies (in %) for 6×6 incomplete matrices: summer houses

number of matrix elements / order	5	6	7	8	9	10	11	12	13	14	15
sequential	0	0.96	1.82	3.71	4.74	5.66	6.61	7.33	8.35	9.21	10.75
random	0	1.38	2.77	3.49	4.42	4.97	6.25	6.91	8.17	8.19	9.47
Ross	0	1.37	2.50	3.84	4.93	5.45	6.27	7.24	7.85	9.52	10.63

Table 8 The average of CR inconsistencies (in %) for 6×6 incomplete matrices: maps

order \ number of matrix elements	5	6	7	8	9	10	11	12	13	14	15
sequential	0	0.13	0.18	0.25	0.32	0.40	0.48	0.55	0.64	0.72	0.81
random	0	0.06	0.11	0.20	0.40	0.50	0.57	0.65	0.71	0.72	0.80
Ross	0	0.07	0.14	0.23	0.31	0.37	0.51	0.69	0.73	0.83	0.88

Table 7 and *Table 8* suggest that from 0 to the final CR value the averages for both types show an almost linear increase as the sequence progresses. To examine the consistency of the CR index during completion, we performed a mixed-effects linear regression of partial CR index on the fixed terms shown in *Table 9*, and also random, per-subject slope and intercept terms, to account for the correlations within each subject's responses and model natural variation between subjects. A likelihood ratio chi-squared test of the significance of the random effects shows them to be significant (chi-squared(3) = 2895; p-value < .001). *Table 9* shows that as subjects progress through the sequence, the CR index does increase linearly (visual inspection of the data hinted at a possible quadratic component, but this was not statistically significant). This positive association between sequence number and partial CR index is even greater in the summer house task, but it is somewhat dampened by sequential ordering or as matrix dimension increases. We can also see that random ordering gave greater partial CR indices in the map task, but this order effect was largely absent in the summer house task. And, as expected, larger matrices lead to higher partial CR indices – but, as above, this effect is less for the summer house task.

Table 9 Parameter estimates for mixed effect linear regression of partial CR index

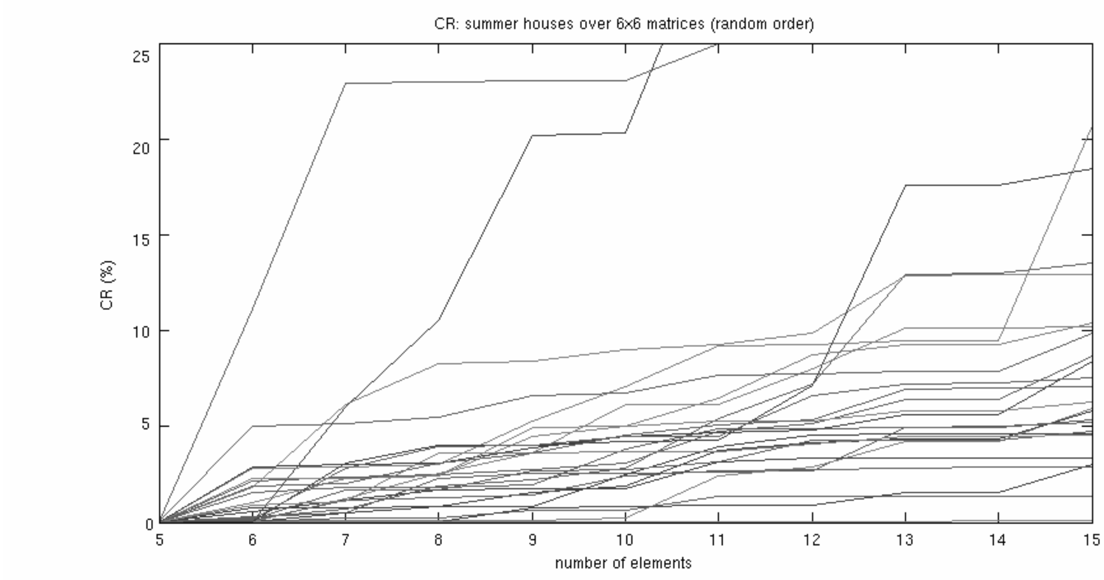
	Estimate	Std error	t-value (df)	p-value
sequence number	.02	.00	10.23 (5463)	.00
sequence number squared	0.00	0.00	.32 (5463)	.75
type				
map	0	N/A	N/A	N/A
summer house	.07	.00	15.63 (5463)	.00
order				
random	0	N/A	N/A	N/A
sequential	-.001	.00	-2.71 (5463)	.01
Ross	-.01	.00	-5.12 (5463)	.00
size	.01	.00	7.04 (5463)	.00
sequence number×type:summer house	.001	.00	39.74 (5463)	.00
sequence number×order:Ross	.00	.00	.86 (5463)	.40
sequence number×order:sequential	-.00	.00	-2.77 (5463)	.01
type:summer house×order:Ross	.02	.00	7.08 (5463)	.00
type:summer house×order:sequential	.00	.00	3.50 (5463)	.00
sequence number×dimension	-.00	.00	-8.66 (5463)	.00
type:summer house×size	-.02	.00	-20.51 (5463)	.00

Overall likelihood ratio chi-sq test for the significance of the random effects gave chi-sq(3) = = 2895, p-value < 0.001

To gain further visual insight *Figure 2* presents individual behavioral inconsistency by showing individual CR values for the random questioning order. All polylines in the figure denote one subject in the group. If we draw a vertical line at every step (from 5 to 15) we see the individual values of the CR inconsistency after having answered that number of questions. Since there were 26 subjects in this session 26 polylines can be seen in the figure representing all of their answers. This figure gives details what lies behind the averages represented in the second row of *Table 3*.

Due to the limited space we could not publish all tables and figures in our analysis of incomplete matrices. Calculations for other matrix sizes than in *Table 7* and *Table 8*, and figures similar to *Figure 2* can be seen on our webpage. (Find it on: http://www.sztaki.mta.hu/~bozoki/BozokiDezsoPoeszTemesi2013).

Figure 2 The individual CR inconsistencies for 6×6 matrices, random questioning order



Yet another way to test the behavioral inconsistency is to trace the score vectors and the corresponding ordering as completion progresses (i.e. using the subsequent incomplete matrices).

First we calculated the score vectors for each level of completion in the sequence and compared them to the final score vector calculated from the corresponding complete PC matrix. Two principles were applied for the comparison of two vectors: cardinal and ordinal. Cardinal view (*Table 10 and Table 11*) treats score vectors as elements of the *n* dimensional Euclidean space, where closeness or similarity of two vectors can be measured by, e.g., Euclidean distance. Ordinal view (*Table 12 and Table 13*) takes only ranks into consideration.

The averages (and our detailed calculations for each subject as it can be checked on our referred webpage) confirmed that the distance of step by step scores from the final score monotonically decreased in each step with a very few exceptions. *Table 10* and *Table 11* include the averages of the Euclidean distance of the score vectors calculated by the Eigenvector Method from the 8×8 complete and incomplete matrices in case of both tasks broken down by questioning order.

Table 10 Euclidean distance from the final scores for 8×8 incomplete matrices: summer houses

order \ number of matrix elements	7	8	9	10	11	12	13	14	15	16	17
sequential	0.17	0.16	0.16	0.15	0.13	0.12	0.11	0.10	0.09	0.08	0.06
random	0.22	0.22	0.20	0.20	0.18	0.17	0.13	0.11	0.10	0.10	0.09
Ross	0.23	0.20	0.16	0.15	0.13	0.13	0.12	0.11	0.10	0.09	0.08

order \ number of matrix elements	18	19	20	21	22	23	24	25	26	27	28
sequential	0.06	0.06	0.05	0.04	0.04	0.03	0.03	0.03	0.02	0.01	0.00
random	0.08	0.08	0.06	0.06	0.05	0.05	0.05	0.03	0.03	0.02	0.00
Ross	0.07	0.07	0.06	0.05	0.05	0.04	0.03	0.03	0.02	0.02	0.00

Table 11 Euclidean distance from the final scores for 8×8 incomplete matrices: maps

order \ number of matrix elements	7	8	9	10	11	12	13	14	15	16	17
sequential	0.054	0.054	0.047	0.044	0.043	0.040	0.039	0.035	0.033	0.031	0.030
random	0.059	0.062	0.057	0.048	0.045	0.046	0.044	0.035	0.033	0.032	0.029
Ross	0.091	0.081	0.053	0.048	0.042	0.033	0.032	0.030	0.028	0.026	0.025

order \ number of matrix elements	18	19	20	21	22	23	24	25	26	27	28
sequential	0.029	0.021	0.015	0.015	0.010	0.010	0.007	0.006	0.003	0.003	0.000
random	0.028	0.023	0.020	0.017	0.015	0.010	0.010	0.008	0.008	0.006	0.000
Ross	0.024	0.021	0.017	0.016	0.015	0.014	0.011	0.009	0.009	0.005	0.000

Again these tables show a robust impact of the problem type. Nevertheless, it seems that the questioning order does not have an impact, although one may point out that most of the maximum distance values are found in the rows of the random method. Our conclusion can be summarized in the next statement.

Proposition 4: Throughout the pairwise comparison procedure the majority of the decision makers display a quasi-consistent behavior.

As a next step we calculated the resulted order of alternatives in each step for each subject and for every task, and correlated these values with the final order. *Table 12* summarizes the Spearman coefficients for the 6×6 matrices. The coefficient gives +1 if the ranks are identical, and gives –1 if they are totally reversed.

Table 12 Spearman rank correlation coefficients for 6×6 incomplete matrices

type \ number of matrix elements	5	6	7	8	9	10	11	12	13	14	15
summer houses	0.82	0.88	0.90	0.92	0.93	0.94	0.96	0.97	0.97	0.98	1.00
maps	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

The high coefficients indicate that the orders are very close to the final orders for the map problem from the very initial steps. When we looked at the individual charts which have been generated from our database, we found, indeed, that only a very few of the 26 subjects had a different ordering compared to the final one.

The correlations reported in *Table 12* gave us an overall insight about the similarity of orders. (More individual charts and rank correlation tables can be found on the webpage referred on page 13). Now, we can ask that in a particular questioning process according to the final order how many matrix elements are on the right place after a certain number of pairwise comparisons are completed. Our next research question is: *Can we reduce the number of pairwise comparisons estimating the preferences if they are calculated from incomplete data?*

The cells of the 6×6 matrices in *Table 13* contain the proportion of those matrix elements which are on the proper place according to the final order. Four matrices were calculated: after 10 and 14 pairwise comparisons for both types of problems.

Table 13 Position of the elements of 6×6 complete PC matrices after 10 and 14 comparisons in %

Number of comparisons: 10

	1	2	3	4	5	6
1	99	1	0	0	0	0
2	1	99	0	0	0	0
3	0	0	99	1	0	0
4	0	0	1	99	0	0
5	0	0	0	0	99	1
6	0	0	0	0	1	99

a) map

Number of comparisons: 14

	1	2	3	4	5	6
1	100	0	0	0	0	0
2	0	100	0	0	0	0
3	0	0	100	0	0	0
4	0	0	0	100	0	0
5	0	0	0	0	100	0
6	0	0	0	0	0	100

b) map

Number of comparisons: 10

	1	2	3	4	5	6
1	83	16	0	1	0	0
2	15	65	18	1	0	1
3	1	17	73	9	0	0
4	0	2	8	81	8	1
5	1	0	1	8	73	17
6	0	0	0	0	19	81

c) summer house

Number of comparisons: 14

	1	2	3	4	5	6
1	92	7	1	0	0	0
2	7	84	8	0	0	1
3	1	9	90	0	0	0
4	0	0	1	98	1	0
5	0	0	0	2	95	3
6	0	0	0	0	4	96

d) summer house

Consider a 6×6 incomplete matrix of a map task after 10 comparisons. The elements of the first row of *Table 13a* show that 99% of the alternatives which should be on the first place (according to the known final order) are in fact on the first place after completing 10 pairwise comparisons, and only 1% is on the second place. From the second row we can see that 99% of the alternatives which should be on the second place are in fact in this position after 10 questions, again. The interpretation of *Table 13c* is similar: e.g., we can see from the first row that 83% of the alternatives which should be on the first place are in fact on the first place after completing 10 comparisons, 16% are on the second and 1% is on the fourth place. We can draw the conclusion that for the summer houses task the fit after 10 pairwise comparisons is not as good as for the map tasks. After 14 comparisons (only the last comparison is missing in *Table 13b* and *13d*) both tables demonstrate a good fit. Note, that the fitness for the map task is perfect, while the summer houses tasks need the last comparison much more. Tables of different matrix sizes and various numbers also confirm our findings.

Table 13 suggests that we can use the incomplete matrices for approximation purposes. Executing the entire questioning procedure seems to be unnecessary: we are able to receive a fairly good estimation of the scores and/or rankings after having a certain number of pairwise comparisons completed. In addition we have also learnt that for objective tasks a significantly fewer pairwise comparisons are required to obtain a good estimation than for subjective tasks. This finding sets clear future research agenda, namely to determine the minimum required number of pairwise comparisons in order to obtain reliable estimations.

Proposition 5: Incomplete PC matrices can be used to approximate the final results of all pairwise comparisons.

Our database allowed us to analyze another characteristic of inconsistency. The intransitive triads (in the ordinal sense, i.e., A > B, B > C, yet, C > A, where > denotes 'is preferred to' or 'is larger than' etc., also analyzed by *Gass (1998)* and *Kéri (2011))* provided information about structural problems with a PC matrix. The PC matrix cannot be consistent if there is at least one intransitive triad present among the comparisons. From the viewpoint of CM we can easily identify the inconsistent decision makers, as they have more than one inconsistent triad. Note, that from a CR viewpoint a CR value which is greater than 10% leads to similar conclusion: subjects featured by high CR are not accepted in a decision making process.

Number of intransitive triads can be compared to at least two values. One is total number of triads, that is $n(n-1)(n-2)/6$. Another and more relevant benchmark is the maximal number of intransitive triads in a pairwise comparison matrix of the same size. *Kendall and Smith (1940)* proved that the maximal number of intransitive triads is $(n^3 - n)/24$ if n is odd and $(n^3 - 4n)/24$ if n is even, that is 2, 8 and 20 as $n = 4, 6, 8$, respectively. At the moment we did not investigate the properties of a new index generated by one of these values.

From our 454 complete PC matrices 38 had at least one intransitive triad. Out of these thirty-eight 34 was among the subjective type problems. One of these 38 was among the 4×4 size matrices, 7 of the 38 were among the 6×6 size matrices, and 30 of the 38 were among the 8×8 size matrices. These facts are in accordance with *Proposition 1* and with *Proposition 2*. Out of the same thirty-eight intransitive triads 13 were from the sequential questioning order, 11 matrices were from the random order and 14 were from Ross order. This is in line with *Proposition 3.*

Furthermore, we assigned CR values to each of these thirty-eight matrices. *Table 14* shows that most of the matrices with high numbers of intransitive triads (from 3 to 7) have CR values above the 10% threshold – as we expected. On the other hand, ten of the PC matrices with CR values below the 10% threshold have 1 or 2 intransitive triads. It could be in the interest of our further research to analyze the properties of these matrices.

Table 14 CR values and the number of intransitive triads

CR in %	number of intransitive triads							total
	1	2	3	4	5	6	7	
0 - 5	3	1						4
5 - 10	4	1			1			6
10 - 20	9	2		2				13
20 - 40	4	1	2	1		3		11
40 -		2				1	1	4
total	20	7	2	3	1	4	1	38

As we mentioned in the introduction highly inconsistent matrices can be corrected. Intransitive triads indicate that the source of inconsistency is the presence of one or more outliers. *Bozóki, Fülöp and Poesz (2011)* proposed a procedure to make these matrices consistent by changing their element(s) and they also determined the number of elements that are necessary to be modified to eliminate inconsistency. In our current database there are 47 of those matrices in which inconsistency can be eliminated by modifying maximum 2 elements (5 of them was consistent without any changes). However, "eliminating the inconsistency" may potentially lead to controversial results, e.g. it is possible that the priorities at the end of this process would distract the DM from the real priorities. One has to be cautious in applying correction methods without confirming its results with the DM. An "automatic" execution of elimination can change the real goal of the decision maker and/or change his real preferences. Thus, correction methods can only be cautiously applied and only with the approval of the decision maker.

4 Conclusions

We built a database of empirical PC matrices to test simultaneously the impact of the nature (i.e., subjective versus objective tasks) of the MADM problem, the size of the PC matrix, and the questioning order. Another goal of our study was to extend the investigation for incomplete PC matrices. However, we make no claim that our sample was representative and the experiment was incentive compatible. Instead, we propose that this experimental design could potentially be used in systematic investigations on how human cognitive capacity interferes with consistent preference ranking.

We found evidence that the nature of the task and the size of the matrices do affect the decision maker's consistency (measured by inconsistency indices). In fact, increasing matrix size for subjective task has more sever impact on consistency. That is increasing the matrix size in a subjective task will lead to greater inconsistency than for objective task. However, we also found that the questioning order does not have an impact on consistency of either tasks.

Based on our findings we suggest that decision-support systems with built-in pairwise comparison questioning modules and built-in inconsistency calculations could be developed in several directions. One of the possible improvements is fine tuning these systems in respect of their contents. Since consistency checks are crucial we recommend a parallel use of various approaches exploring the existence and the source of inconsistency in order to apply a proper correction method, if necessary. Thus correction methods may be reconsidered and may be incorporated into the model. Moreover, we are assuming that consistency could also benefit from increasing the interactivity of the decision-aiding software packages *(Temesi, 2006).*

Furthermore, we also presented preliminary results on incomplete matrices obtained by a step-by-step procedure. We used new techniques to compute the elements of the incomplete matrices to reveal the decision maker's behavior in the progress of completing the matrices. Our results are promising as they hint applicability of incomplete matrices for approximating the complete PC matrices, along with finding thresholds in reducing the number of questions in large-size pairwise comparison problems.

One of our future research directions can be to design and conduct systematic studies to investigate how certain characteristics of matrices (i.e., size, content, importance for the decision maker, relevance to important/unimportant aspects of the decision makers' state) impact behavioral consistency. These characteristics need to go under systematic research since our findings could potentially improve the real-world applications of decision-support systems.

Furthermore, we conjecture that there are also important questions to be answered when designing decision support systems for solving MADM problems. One of the questions to investigate is the differences between paper-based or computer based procedures, or seek out the potential difference between numerical vs. graphical response scales, or study whether one should or should not provide inconsistency feedback during completion, or decide whether offering the corrections to the decision maker could make him behave more consistently.

In addition, our long-term research goal is to explore the application of incomplete matrices. Most importantly, we are planning a systematic investigation on exploring what kind of decision support could make decision makers better off and make them behave more consistent. Based on our results and on other studies we believe that a more in-depth and systematic investigation on inconsistency indices could contribute to achieve that goal.

References

Aguarón, J., Moreno-Jiménez, J.M. (2003): The geometric consistency index: Approximated thresholds, *European Journal of Operational Research* 147, pp. 137–145.

Bana e Costa, C.A., Vansnick, J-C. (2008): A critical analysis of the eigenvalue method used to derive priorities in AHP, *European Journal of Operational Research* 187, pp. 1422-1428.

Bozóki, S., Fülöp, J, Koczkodaj, W.W. (2011): An LP-based inconsistency monitoring of pairwise comparison matrices, *Mathematical and Computer Modelling,* 54(1-2), pp. 789-793.

Bozóki S., Fülöp, J., Poesz, A. (2011): On pairwise comparison matrices that can be made consistent by the modification of a few elements, *Central European Journal of Operations Research*, 19(2), pp. 157-175.

Bozóki, S., Fülöp, J., Rónyai, L. (2010): On optimal completions of incomplete pairwise comparison matrices, *Mathematical and Computer Modelling*, 52, pp. 318-333.

Bozóki, S., Rapcsák, T. (2008): On Saaty's and Koczkodaj's inconsistencies of pairwise comparison matrices, *Journal of Global Optimization*, 42(2), pp. 157-175.

Brunelli, M., Fedrizzi, M. (2011): Characterizing properties for inconsistency indices in the AHP *Proceedings of the International Symposium on the Analytic Hierarchy Process (ISAHP)*, 15-18 June, 2011, Sorrento (Naples), Italy.

Choo, E.U., Wedley, W.C. (2004) A common framework for deriving preference values from pairwise comparison matrices, *Computers & Operations Research* 31, pp. 893–908.

Gass, S. (1998): Tournaments, Transitivity and Pairwise Comparison Matrices, *Journal of the Operational Research Society,* 49, pp.616-624.

Gass, S.I. and Rapcsák, T. (2004): Singular value decomposition in AHP, *European Journal of Operational Research*, 154, pp. 573-584.

Gass, S.I., Standard, S.M. (2002): Characteristics of positive reciprocal matrices in the analytic hierarchy process, *Journal of the Operational Research Society,* 53(12), pp.1385-1389.

Harker, P.T. (1987): Incomplete pairwise comparisons in the analytic hierarchy process, *Mathematical Modelling*, 9(11), pp. 837-848.

Kendall, M.G., Smith, B.B. (1940): On the Method of Paired Comparisons, *Biometrika*, 31(3-4) (March 1940), pp. 324-345.

Kéri, G. (2011). On qualitatively consistent, transitive and contradictory judgment matrices emerging from multiattribute decision procedures, *Central European Journal of Operations Research,* 19, pp. 215-224.

Koczkodaj, W.W. (1993): A new definition of inconsistency of pairwise comparisons, *Mathematical and Computer Modelling* 8, pp. 79-84.

Koczkodaj, W.W., Herman, M.W., & Orlowski, M. (1997): Using consistency-driven pairwise comparisons in knowledge-based systems. In: *Proceedings of the Sixth International Conference on Information and Knowledge Management,* ACM Press, pp. 91-96.

Linares, P. (2009): Are inconsistent decisions better? An experiment with pairwise comparisons, *European Journal of Operational Research,* 193, pp. 492–498.

Miller, G. A. (1956): The magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, *The Psychological Review,* 63, pp. 81-97.

Murphy, C.K. (1993): Limits on the Analytic Hierarchy Process from its inconsistency index, *European Journal of Operational Research*, 65, pp. 138-139.

Peláez, J.I., Lamata, M.T. (2003): A new measure of consistency for positive reciprocal matrices, *Computers and Mathematics with Applications* 46, pp. 1839–1845.

Poesz, A. (2008): Inconsistency analysis of empirical pairwise comparison matrices, Master's Thesis, Corvinus University of Budapest (in Hungarian)

Ross, R.T. (1934): Optimum orders for the presentation of pairs in the method of paired comparison, *Journal of Educational Psychology,* 25, pp. 375-382.

Saaty, T.L. (1980): *The analytic hierarchy process*, McGraw Hill, New York

Saaty, T.L. and Ozdemir, M.S. (2003): Why the magic number seven plus or minus two, *Mathematical and Computer Modelling*, 38(3-4), pp. 233-244.

Saaty, T.L. (2005): The analytic hierarchy and analytic network processes for the measurement of intangible criteria and for decision-making, In Figueira, J., Greco, S., Ehrgott, M. (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer, New York, pp. 345–407.

Temesi, J. (2006): Consistency of the decision-maker in pairwise comparisons, *International Journal of Management and Decision Making,* 7(2-3), pp. 267-274.

Temesi, J. (2011): Pairwise comparison matrices and the error-free property of the decision maker, *Central European Journal of Operational Research*, 19 (2), pp. 239-249.

Xu, Z., Wei, C. (1999): A consistency improving method in the analytic hierarchy process, *European Journal of Operational Research,* 116, pp. 443-449.