

**Anna PRUSAK<sup>1</sup>**  
**Piotr STEFANÓW<sup>2</sup>**  
**Jacek STROJNY<sup>3</sup>**  
**Monica GARCIA-MELON<sup>4</sup>**

## **THE INFLUENCE OF THE FORM OF THE 9-POINT SCALE IN THE AHP METHOD ON THE CONSISTENCY OF JUDGMENTS**

Quality of decisions depends largely on the ability to correctly define and assess the problem. In the case of complex issues, it is recommended to use decision support methods, e.g. multicriteria methods. The objective of this paper is to report the studies related to the influence of the graphic form of the 9-point, fundamental Saaty's comparison scale used in the AHP method on the consistency of judgments, that is, to measure the fraction of judgments with  $CR > 0.10$ . Since the AHP is one of the most frequently used decision support methods in management, in terms of modeling decision problems, there is a need to explore one of its most frequently discussed problems – inconsistency of results. It will improve the quality of decisions made with the use of this tool. The empirical study was conducted among 540 respondents, using the AHP method. Due to errors in filling in the questionnaire, only 424 questionnaires were included in further analysis. Individual model was prepared and analyzed for each respondent. All results were then entered into a spreadsheet and subjected to statistical analysis. It examines four most commonly used graphic forms of scale (specifically: numerical, two-stage tabular verbal, tabular verbal horizontal and vertical). Chi-square test and F-test showed no significant difference between them in relation to the existence of inconsistent results ( $CR > 0.1$ ). It allows suggest that the form of graphic scale does not affect the consistency of answers. However, additional analysis showed that it affects errors in questionnaires.

**Keywords:** analytical hierarchical process, AHP, consistency, CR.

### **1. INTRODUCTION**

The quality of decision-making processes is a very important factor in the competitiveness of modern organizations. Decision making is in fact procedural and technological feature of the management process<sup>5</sup>. Modeling and optimization of decision-making processes is a frequently discussed research problem in management sciences. As a result,

---

<sup>1</sup> Dr Anna Prusak, Wydział Towaroznawstwa, Uniwersytet Ekonomiczny w Krakowie, e-mail: anna.prusak@uek.krakow.pl

<sup>2</sup> Dr Piotr Stefanów, Wydział Zarządzania i Komunikacji Społecznej, Krakowska Akademia im. A.F. Modrzewskiego, e-mail: pstefanow@afm.edu.pl

<sup>3</sup> Dr Jacek Strojny, Wydział Zarządzania, Politechnika Rzeszowska, e-mail: jstrojny@prz.ed.pl (Author for correspondence)

<sup>4</sup> Dr Monica Garcia-Melon, School of Industrial Engineering, Universitat Politècnica de València, mgarciam@dpi.upv.es

<sup>5</sup> J. Targalski, *Podjęmowanie decyzji* [w:] *Organizacja i zarządzanie*, red. A. Stabryła, J. Trzcieniecki, Warszawa 1986, p. 194.

concepts, methods and related tools have been created, with potential to improve the decisions. Such measures are particularly important when decision problem is complex, requires consideration of many aspects and selection of specific priorities. In such a situation, it is convenient to use the multi-criteria decision support methods, which include, inter alia, analytic hierarchy and network processes (AHP/ANP), discussed in this paper.

The analytic hierarchy and network processes (known as the AHP/ANP), were developed in the 70's by the American mathematician T.L. Saaty and can be considered as the most popular multi-criteria decision support tools (MCDA – Multiple-Criteria Decision Analysis). Their attractiveness is associated with the possibility to use for solving complex organizational problems. This results in a huge number of references in international journal databases<sup>6</sup>. Due to versatility, flexibility and simplicity of these methods, they are used in various fields of science<sup>7</sup>, as well as business practice<sup>8</sup>. There is also a dedicated software for calculations of very complex models – Super Decisions<sup>9</sup>.

Both the AHP and ANP methods are based on the same mathematical assumptions, and the difference between them is due to construction and interpretation of the model. In case of the AHP it is hierarchical structure, in which each element has its precise place (decision goal, criteria, subcriteria, and variants also called „alternatives”)<sup>10</sup>. The structure of relationships between groups of elements of the hierarchy (in the AHP method) implies dependence of the goal of the criteria, the criteria of subcriteria, and in the case of variants it is important to define the extent to which they meet each individual subcriterion. Network models (in the ANP method) allow to consider much more complex and multidirectional relationships<sup>11</sup>. In order to simplify the research procedure described in this paper, the study has been limited only to the AHP method, although the results and conclusions can be considered universal for both tools.

Popularity of the AHP/ANP methods is also related to the fact that they have been thoroughly studied by teams from different countries, both in terms of application and methodological aspects<sup>12</sup>. One of the most frequently studied and discussed areas of the AHP/ANP is consistency (compatibility, logic) of the results (judgments). Consistency in the case of these methods must be seen primarily in mathematical terms, in the context of its specific measure called *Consistency Ratio* (known as CR).

If  $CR > 0.10$ , the procedure requires rejection of repeating all judgments, for which CR exceeded the accepted level. This in practice leads to the loss of lots of data, or is associ-

---

<sup>6</sup> P. Stefanów, A. Prusak, *Badanie wiarygodności i skuteczności skali porównań Saaty'ego w metodzie AHP i ANP* [w:] *Przedsiębiorcze aspekty rozwoju organizacji i biznesu*, red. A. Chodyński, Oficyna Wydawnicza AFM, Kraków 2011, s. 94.

<sup>7</sup> O.S. Vaidya, S. Kumar, *Analytic hierarchy process: An overview of applications*, „European Journal of Operational Research” 2006, Vol. 169, No. 1, pp. 1–2.

<sup>8</sup> T.L. Saaty, *Relative measurement and its generalization in decision making. Why Pairwise comparisons are central in mathematics for the measurement of intangible factors*, The Analytic Hierarchy/Network Process, „Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matematicas” 2008, Vol. 102, No. 2, p. 253.

<sup>9</sup> [www.superdecisions.com](http://www.superdecisions.com)

<sup>10</sup> This is the example of a very basic, four-level model. In fact, there can be more levels.

<sup>11</sup> A. Prusak, P. Stefanów, *AHP – analityczny proces hierarchiczny. Budowa i analiza modeli decyzyjnych krok po kroku*, C.H. Beck, Warszawa 2014, p. 40.

<sup>12</sup> A. Prusak, P. Stefanów, *Badania nad właściwościami metody AHP*, *Folia Oeconomica Cracoviensia*, 2011, nr LII, p. 95.

ated with high costs of repetition of surveys, which in many cases is not possible<sup>13</sup>. For this reason, researchers from all over the world try to develop methods for inducing the expression of more consistent judgments, or computer algorithms which allow for automatic reduction of inconsistencies in the matrix (reduction the value of CR). The latter, however, cause disruption of the original data presented in pairwise comparison matrix. Thus, Gastes & Gaul<sup>14</sup> formulated the following conclusion: „It may be an important topic for future research to understand, why decision makers do not state sufficiently consistent preferences and to control consistency right from the beginning of pairwise comparisons tasks. This should be done without forcing decision makers to erroneous preference statements just because they have to fulfill consistency constraints. One objective may be, to find processes of preference interrogation, which result in more consistent comparison matrices than traditional questioning”.

The study reported in this paper partly fill in this research gap. The objective was to determine to what extent the graphic form of the questionnaire (based on the Saaty's 9-point scale) influence fraction of inconsistent judgments, as measured by CR. The research involved N=540 respondents, which were students participating in lectures on *Statistics* and *Multi-criteria decision-making methods*. The hierarchical model was based on mobile phone, which evaluated according to 4 criteria and presented using 4 different graphic forms of the 9-point scale. Responses (in the form of pairwise comparisons) were introduced to the Super Decisions software – the models were analyzed by each respondent individually. To determine the fraction of matrices within certain consistency levels (CR), the results obtained for each respondent in Super Decisions (weights and CR for each form of the scale) were then put to a spread sheet.

The first part of this paper is a review of the existing knowledge on the graphic form of the AHP questionnaire. The second part presents stages and discusses the sources and consequences of inconsistency of judgments. The next part contains a description of methodology, and the following one is a discussion of the results. Finally, the conclusions were formulated with reference to the areas requiring further research.

## 2. OVERVIEW OF THE EXISTING RESEARCH ON THE GRAPHIC FORM OF THE AHP QUESTIONNAIRE

The influence of the graphic form of the AHP questionnaire on decision makers' judgments is a rarely discussed topic in the literature. It should be emphasized that AHP is one of the most popular decision support methods, widely used both in science and in practice. Therefore, it is particularly important to look more closely at its data collection instruments. The possible errors arise mainly at the stage of data collection, and their cause is primarily the human factor: fatigue because of too many comparisons of different pairs of repeating elements, or rush in expressing judgments, leading to their randomness.

---

<sup>13</sup> B. Apostolou, J.M. Hassell, *An empirical examination of the sensitivity of the analytic hierarchy process to departures from recommended consistency ratios*, "Mathematical and Computer Modelling" 1993, Vol. 17, No. 4-5, p. 169.

<sup>14</sup> D. Gastes, W. Gaul, *The Consistency Adjustment Problem of AHP Pairwise Comparison Matrices* [in:] A. Diamantopoulos, W. Fritz, L. Hildebrandt (eds.), *Quantitative Marketing and Marketing Management*, Gabler Verlag, Wiesbaden 2012, p. 61.

Meanwhile, the only guideline for this step of the AHP/ANP is to use verbal scale instead of the numeric one<sup>15</sup>.

In the present study, it has been assumed that the graphic form of the AHP questionnaire may induce or reduce the correctness and consistency of judgments. It is commonly known that the transparency of the research questionnaire eliminates mistakes made by the respondents and increase their intrinsic motivation to answer conscientiously. The influence of various aspects of the appearance of the questionnaire on the responses provided was studied in the area of marketing, as it is not a subject specific to the AHP method. For example, Preston & Colman<sup>16</sup> and Weathers et al.<sup>17</sup> studied the influence of a number of degrees of the scale on the correctness and reliability of responses. Weijters et al.<sup>18</sup> found a strong correlation between the response style and the format of the presented scale.

In the case of the graphic form of the AHP questionnaire, certain arrangements of pairwise comparisons may increase readability, and thus reduce the risk of errors and inconsistencies (or induce thereof). These relationships have not been so far described in the literature, despite the fact that the Consistency Ratio CR (discussed in more detail in the next section) is one of the most frequently investigated aspects of the AHP<sup>19</sup>.

However, other features of the AHP questionnaire have been studied. For example, Webber et al.<sup>20</sup> conducted an experiment on the effect of the graphical presentation of pairwise comparison scale and the order of the questions on the final results. It showed the existence of a link between the reporting format of the questionnaires and values of priorities. A weak relationship between priorities and type of the scale used (numerical, verbal, graphical) was also identified. The results for which  $CR > 0,10$  were excluded from the analysis and the problem of consistency was not considered here.

### 3. SOURCES AND CONSEQUENCE OF INCONSISTENCY OF JUDGMENTS

The ability to calculate a ratio measuring consistency of judgments is a consequence of the specificity of the AHP. Its stages were described in many publications<sup>21</sup>. They can be described as follows: 1) construction of the decision-making hierarchical model, 2) preparing the research questionnaire based on the 9-point comparison scale, 3) data gathering (expressing judgments by pairwise comparisons), 4) calculating weights (priorities), 5) calculating the consistency of judgments using CR (it should not exceed 0.10), 6) aggregating the results from different respondents. After building the hierarchical model one

<sup>15</sup> A. Prusak, P. Stefanów, *AHP – analityczny...*, p. 91.

<sup>16</sup> C.C. Preston, A.M. Colman, *Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences*, "Acta Psychologica" 2000, Vol. 104, No. 1.

<sup>17</sup> D. Weathers, S. Sharma, R.W. Niedrich, *The Impact of the Number of Scale Points, Dispositional Factors, and the Status Quo Decision Heuristic on Scale Reliability and Response Accuracy*, "Journal of Business Research" 2005, Vol. 58, pp. 1516–1524.

<sup>18</sup> B. Weijters, E. Cabooter, N. Schillewaert, *The effect of rating scale format on response styles: The number of response categories and response category labels*, "International Journal of Research in Marketing" 2010, Vol. 27, pp. 236–247.

<sup>19</sup> A. Prusak, P. Stefanów, *Badania nad właściwościami...*, p. 95.

<sup>20</sup> S.A. Webber, B. Apostolou, J.M. Hassel, *The sensitivity of the analytic hierarchy process to alternative scale and cue presentations*, "European Journal of Operational Research" 1996, Vol. 96, No. 2.

<sup>21</sup> A. Prusak, P. Stefanów, *AHP – analityczny...*; [6] Ishizaka H., Labib A., *Analytic Hierarchy Process and Expert Choice: Benefits and Limitations*, "ORInsight" 2009, Vol. 22, No. 4, pp. 201–220.

needs to obtain the source data (judgments) necessary to determine weights (priorities). The judgments are expressed by decision makers (experts) as pairwise comparisons of the elements within the respective groups (clusters) of the hierarchical model. They reflect opinions, knowledge and feelings of the experts on the analyzed "fragment" of the decision problem. Comparisons are designed to express dominance (advantage) of one element over the other. The „dominance” can mean „importance”, „significance”, „probability”, „preference”, „relevance” and other relationships, which are defined depending on the analyzed decision problem, type of elements and their place in the hierarchical structure.

Saaty proposed a special bipolar 9-point comparison scale, known as the fundamental scale. It has in fact 17 degrees (from „1” to „9” on each side) and it is in fact the only scale used in the AHP. This scale is also built-in dedicated software. If one uses a verbal scale, the indicated degree of dominance must be transposed into numerical values as follows (A and B are the elements being compared):

- 1) A and B have equal importance („1”, middle of the scale),
- 2) A is slightly more important than B or B is slightly more important than A („3” on the left or the right side of the scale),
- 3) A is moderately more important than B, or B is moderately more important than A („5”),
- 4) A is much more important than B, or B is much more important than A („7”),
- 5) A is extremely more important than B, or B is entirely more important than A („9”).

If the respondent's preferences lie somewhere between the main categories („1”, „3”, „5”, „7”, „9”), they are represented by even numbers („2”, „4”, „6”, „8”). In case of verbal scale, the respondents are asked to express their indecision by indicating judgments on the verge of the main categories. These numbers are then introduced to a pairwise comparison  $n \times n$  matrix A, based on which priority weight are calculated.

It has been repeatedly emphasized in the literature that the stage of data collection, or making comparisons, generates the most problems<sup>22</sup>. This is so for two reasons. First, the comparisons (based on which priority weights are calculated) are based on the judgments expressed by individual experts, which by their nature can be subjective. However, this is a property common to all social research. Secondly, one of the principles of obtaining the pairwise comparison is taking into account all possible combinations of pairs of elements of the model (each element must be compared with each other located in the same group of hierarchical structure). It causes their redundancy in relation to the number of comparisons necessary for calculating priorities. This redundancy is the basis for measuring the degree of consistency of judgments, using the following formula:

$$CR = \frac{CI}{RI},$$

where:

CI – Consistency Index, which is calculated as follows:

<sup>22</sup> T.L. Saaty, K. Peniwati, *Group Decision Making: Drawing out and Reconciling Differences*, RWS Publications, Pittsburgh 2007.

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

where:

$\lambda_{\max}$  – principal eigenvalue

n – number of elements compared.

RI – Random Index, an average CI of randomly chosen matrices (Table 1).

Table 1. Random Indices for  $n$  elements based on simulation of 100000–500000 matrices

n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10	n=11	n=12	n=13	n=14	n=15
0,5 2	0,8 8	1,1 1	1,2 5	1,3 4	1,4 1	1,4 5	1,49	1,51	1,54	1,56	1,57	1,58

Source: based on J.A. Alonso, M.T. Lamata, *Consistency in the Analytic Hierarchy Process – A New Approach*, “International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems” 2006, Vol. 14, No. 4.

The number of pairs of elements (N) for each group is presented by the following formula:

$$N = \frac{k(k-1)}{2}$$

where  $k$  is the number of elements compared within the same group.

According to the formula, a group composed of elements  $k=3$  requires the formation of  $N=3$  pairs, while group consisting of 5 elements requires 10 pairs, 7 elements – 21 pairs, and 9 elements – 45 pairs. Therefore, it is recommended that the number of elements in the group should not exceed 9 elements, and preferably have 7, which is referred to as the magic number  $7+/-2$ <sup>23</sup>. It indicates that the greater the number of elements in the group, the greater the risk of inconsistency when making judgments. Respondent, weary of having to make a large number of comparisons of the same elements in different configurations, can provide a completely random responses, which ultimately lead to inconsistent results<sup>24</sup>.

Any comparison made randomly generates inconsistencies, increasing the value of CR, which cannot exceed 0,10 (10%). The AHP/ANP methods are very sensitive even to small inconsistencies, so they have many critics, who consider it the greatest disadvantage of these tools. Moreover, high value of CR is not always due to the respondent, but may result from a lack of homogeneity of the elements. An example of such a situation has been provided in: Stefanów & Prusak<sup>25</sup>. Criticism of CR has been expressed in the literature with respect to the use of this particular ratio as the measure of the degree of consis-

<sup>23</sup> G.A. Miller, *The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information*, “Psychological Review” 1956, Vol. 63, No. 2; T.L. Saaty, M.S. Ozdemir, *Why the magic number seven plus or minus two*, “Mathematical and Computer Modelling” 2003, Vol. 38, No. 3–4, pp. 233–244.

<sup>24</sup> A. Prusak, P. Stefanów, *AHP – analityczny...*, s. 250.

<sup>25</sup> P. Stefanów, A. Prusak, *Badanie wiarygodności...*, p. 287.

tency<sup>26</sup>, as well as to the fact that the acceptable level of CR ( $\leq 0,10$ ) is too restrictive<sup>27</sup>. However, the possibility to calculate CR is perceived an advantage over other decision support methods, which do not provide any control over consistency of opinions. Thus, one cannot objectively say how much the respondent was involved in decision making process. In the case of AHP/ANP, it is easy to check whether comparisons are random.

Inconsistent matrices, where the value of CR exceeds 0,10 should be considered as having low informational value, and it is necessary to repeat the analysis<sup>28</sup>. However, experience has shown that inconsistencies significantly exceeding the level of CR=0,10 take place even for matrices with a small number of elements<sup>29</sup>. Moreover, the study by Apostolou & Hassel<sup>30</sup> revealed no significant differences between priority weights for which CR $\leq 0,10$  and those with CR $>0,10$ . Therefore, they postulated not to reject too rashly the matrices, which are inconsistent according to the Saaty's criterion. It was met with criticism questioning i.a. method of data collection for this study – remotely, via e-mail rather than individual, moderated session<sup>31</sup>.

Minimizing CR should not be the purpose in itself, yet it is an important indicator of the quality of the results. There are various ways to reduce inconsistency (CR), for example, using mathematical algorithms. However, such procedures may lead to perturbation of original matrices<sup>32</sup>. The only way to avoid such problem is to take appropriate action at the stage of data collection process, which involves construction of the AHP questionnaire.

#### 4. METHODOLOGY

The study aimed to answer the following research question: does the graphical presentation of the 9-point comparison scale affect the consistency of judgments (as measured by CR)? It involved N=540 respondents – students participating in lectures on *Statistics* and *Multi-criteria decision-making methods* at the A.F. Modrzewski Krakow University (Faculty of Management and Social Communication and the Faculty of Law, Administration and International Relations), as well as the Cracow University of Economics (Department of Commodity Sciences). The study was conducted from November to December 2013. Similar group (consisting of 340 students) was asked to participate in the research by Webber et al.<sup>33</sup>.

The group was considered expert for this experiment because: 1) the information was universal in nature and objects used in the model can be successfully assessed by the majority of the population (mobile phone); 2) the objective was not to gather the information

<sup>26</sup> J.A. Alonso, M.T. Lamata, *Consistency in the Analytic...*, pp. 445–459.

<sup>27</sup> B. Apostolou, J.M. Hassell, *An empirical examination of the sensitivity of the analytic hierarchy process to departures from recommended consistency ratios*, "Mathematical and Computer Modelling" 1993, Vol. 17, No. 4–5, pp. 163–170.

<sup>28</sup> T.L. Saaty, L.G. Vargas, *The Logic of Priorities*, Kluwer Nijhoff Publishing, Massachusetts 1982.

<sup>29</sup> A. Prusak, P. Stefanów, *Badania nad właściwościami...*

<sup>30</sup> B. Apostolou, J.M. Hassell, *An empirical...*

<sup>31</sup> P. Chu, J.K. Liu, *Note on consistency ratio*, "Mathematical and Computer Modelling" 2002, Vol. 35, No. 9–10, pp. 1078.

<sup>32</sup> J.S. Finan, W.J. Hurley, *The analytic hierarchy process: Does adjusting a pairwise comparison matrix to improve the consistency ratio help?*, "Computers & Operations Research" 1997, Vol. 24, No. 8, pp. 749.

<sup>33</sup> S.A. Webber, B. Apostolou, J.M. Hassel, *The sensitivity...*

about preferences of the functional characteristics of the objects, but to explore a methodological aspect of making pairwise comparisons with the use of various graphical forms of the fundamental scale. Studies were conducted within small groups (sessions), most of them consisting of several respondents (no more than 25 persons). The respondents were asked to identify which characteristics (criterion) of the object is more important and to what extent. The study was based on the AHP evaluation of the mobile phone in terms of: (1) operating time (battery), (2) weight, (3) size of the display, (4) touch screen.

Because of the risks that order of questions will affect the results, the authors decided to randomize the study. For four criteria it gives  $4!$  (factorial), that is 24 possible versions of each questionnaire. For each of the four questionnaires we prepared 24 different versions of the sequence of questions, so the total number of different questionnaires was 96. After receiving the questionnaire and short introduction to the study, respondents were asked to read the purpose of the study. The questionnaire used both numerical and verbal form of the 9-point comparison scale, specifically:

- 1) numerical scale (Fig. 1),
- 2) two-step verbal scale (Fig. 2),
- 3) horizontal tabular verbal scale (Fig. 3),
- 4) vertical tabular verbal scale (Fig. 4).

Fig. 1. Numerical scale

\_\_\_\_\_ **QUALITY** : **COST** \_\_\_\_\_

Source: based on S.A. Webber, B. Apostolou, J.M. Hassel, *The sensitivity...*, p. 361.

Fig. 2. Two-step verbal scale

**Which feature is more important when you buy a car?**      **To what degree?**

**QUALITY** \_\_\_\_\_ or **COST** \_\_\_\_\_

tak samo ważne	średnio	dużo	bardzo dużo	całkowicie
[ ]	[ ]	[ ]	[ ]	[ ]

Source: based on S.A. Webber, B. Apostolou, J.M. Hassel, *The sensitivity...*, p. 361.



Fig. 3. Horizontal tabular verbal scale

criterion	degree of dominance										criterion
	extreme dominance	strong dominance	moderate dominance	weak dominance	equal importance	weak dominance	moderate dominance	strong dominance	extreme dominance		
<b>QUALITY</b>											<b>COST</b>

Source: own research.

Fig. 4. Vertical tabular verbal scale

<p><b>QUALITY (Q)</b></p> <p>Which feature is more important when you buy a car, and how more important?</p>	<input type="checkbox"/>	(Q) is extremely more important than (C)
	<input type="checkbox"/>	(Q) is much more important than (C)
	<input type="checkbox"/>	(Q) is moderately more important than (C)
	<input type="checkbox"/>	(Q) is slightly more important than (C)
	<input type="checkbox"/>	Both are equally important
	<input type="checkbox"/>	(C) is slightly more important than (Q)
	<input type="checkbox"/>	(C) is moderately more important than (Q)
	<input type="checkbox"/>	(C) is much more important than (Q)
	<input type="checkbox"/>	(C) is extremely more important than (Q)
	<b>COST (C)</b>	

Source: own research.

Numerical scale (Fig. 1) requires the respondent to enter a specific number in the blank space. The scale in Fig. 2 was defined as verbal two-step, because the respondent indicates dominance in two steps: 1) which of the two compared elements is more important, 2) how much this dominance is. The breakdown of judgment in two parts helps to concentrate on just one aspect, which in turn may affect the consistency of answers. The horizontal tabular verbal scale (Fig. 3) is the form most commonly used due to space-saving (all comparisons at one table), and may be also in numeric version (numbers instead of verbal expressions). The advantage of vertical tabular verbal scale (Fig. 4) is its transparency, while the disadvantage is that it takes a lot of space and significantly increases the volume of the questionnaire (which can, among others, demotivate the respondents).

Data from the questionnaires was introduced to *Super Decisions* software – each questionnaire required building individual model. Pre-analysis was conducted to check the correctness of filling in questionnaires to be discarded from further studies (eg. mistakes such as lack of response to one question, responses ticked at both sides of the scale). Thus, the number of the valid questionnaires were reduced to N=424. The results (values of the

priority weights and CR for each questionnaire) were then entered into a spread sheet, along with information on the type of the graphic form of the scale. Fractions (proportions) of inconsistent questionnaires (a quotient of the inconsistent questionnaires to all the questionnaires filled in correctly) have been calculated. As a next step, statistical tests were used to verify whether the graphical form of the scale have a significant impact on the value of CR.

## 5. RESULTS

The analysis included only the questionnaires completed correctly. It then examines whether distributions of CR for individual graphical forms of the questionnaire are similar to each other, or are they different (Table 2).

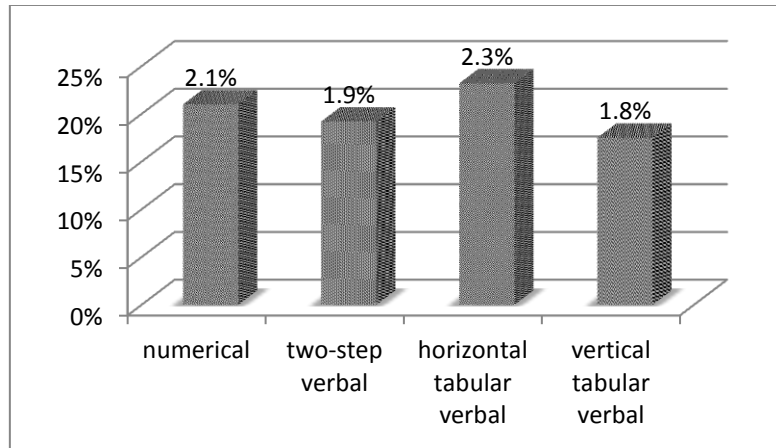
Table 2. Distribution of CR for correctly completed questionnaires (matrices) for various forms of scale

Number of correctly completed questionnaires (N)								
Intervals CR	N=105		N=78		N=99		N=104	
	horizontal tabular verbal		numerical		vertical tabular verbal		two-step verbal	
		%		%		%		%
0,0-0,1	22	21,0%	15	19,2%	23	23,2%	18	17,6%
0,1-0,2	30	28,6%	17	21,8%	24	24,2%	32	31,0%
0,2-0,3	19	18,1%	18	23,1%	16	16,2%	24	23,2%
0,3-0,4	10	9,5%	9	11,5%	12	12,1%	15	14,1%
0,4-0,5	8	7,6%	13	16,7%	9	9,1%	4	3,5%
0,5-0,6	1	1,0%	1	1,3%	4	4,0%	1	0,7%
0,6-0,7	1	1,0%	0	0,0%	1	1,0%	2	2,1%
0,7-0,8	1	1,0%	1	1,3%	3	3,0%	0	0,0%
0,8-0,9	1	1,0%	0	0,0%	1	1,0%	1	1,4%
0,9-1,0	2	1,9%	0	0,0%	1	1,0%	1	1,4%
>1,1	10	9,5%	4	5,1%	5	5,1%	5	4,9%

Source: own research.

The results presented in Table 2 show that the distributions of CR are similar, regardless of what graphic form of the questionnaire was used in the study (Fig. 5).

Fig. 5. Distribution of CR dependent on the scale



Source: own research.

In the next step, it has been verified using chi-square test and F-test<sup>34</sup>.

#### **Chi-square test**

The following hypotheses have been formulated:

- $H_0$ : all fractions are the same (graphic form of scale does not affect the fraction of inconsistent matrices),
- $H_1$ : not all of the fractions are the same (graphic form of scale does not affect the fraction of inconsistent matrices).

First, the expected value was calculated, which together with empirical values are shown in Table 3, and then  $\chi^2$  was set.

Table 3. Empirical and expected values and for chi-square test

Form of scale	Empirical values		N	Expected values		N
	CR>0,10	CR≤0,10		CR>0,10	CR≤0,10	
horizontal tabular verbal scale	83	22	<b>105</b>	83,95	21,05	<b>105</b>
numerical scale	63	15	<b>78</b>	62,36	15,64	<b>78</b>
vertical tabular verbal scale	76	23	<b>99</b>	79,15	19,85	<b>99</b>
two-step verbal scale	117	25	<b>142</b>	113,53	28,47	<b>142</b>
<b>ALL</b>	<b>339</b>	<b>85</b>	<b>424</b>	<b>339</b>	<b>85</b>	<b>424</b>

Source: own research.

<sup>34</sup> J. Steczkowski, M. Woźniak, K. Zajac, A. Zeliaś, *Statystyka matematyczna w zastosowaniach*, Akademia Ekonomiczna w Krakowie, Kraków 1996.

$$\begin{aligned}
 \chi^2 = & \frac{(83 - 83,95)^2}{83,95} + \frac{(22 - 21,05)^2}{21,05} + \frac{(63 - 62,36)^2}{62,36} + \frac{(15 - 15,64)^2}{15,64} + \frac{(76 - 79,15)^2}{79,15} + \frac{(23 - 19,85)^2}{19,85} + \\
 & + \frac{(117 - 113,53)^2}{113,53} + \frac{(25 - 28,47)^2}{28,47} = 1,25
 \end{aligned}$$

For adopted level of significance  $\alpha=0,05$  and for  $s=k-1$  (3) degrees of freedom the critical value was read from the appropriate tables:  $\chi^2_{\alpha}=7,81$ . Because  $\chi^2 < \chi^2_{\alpha}$  ( $1,25 < 7,81$ ), there is no reason to reject the null hypothesis: the influence of the graphical form of the questionnaire on the consistency (CR) cannot be confirmed.

### F-test

As in the case of chi-square, the null hypothesis ( $H_0$ ) states on the equality of all fractions:  $H_0: p_1 = p_2 = \dots = p_k$ ,

where:

$p$  -  $i$ -fraction, for  $k \geq 2$ .

Alternative hypothesis ( $H_1$ ) says that at least two fractions differ significantly:

$H_1$ : not all fractions are equal.

Table 4 presents the summary of calculations.

Table 4. Results of calculations for F-test

Form of scale	N	CR>0,10		$j_i$	$j_i - \bar{j}$	$(j_i - \bar{j})^2$	$n_i(j_i - \bar{j})^2$
			%				
horizontal tabular verbal scale	<b>105</b>	83	79,05%	2,19	-0,0223	0,0005	0,0525
numerical scale	<b>78</b>	63	80,77%	2,23	0,0206	0,0004	0,0312
vertical tabular verbal scale	<b>99</b>	76	76,77%	2,13	-0,0773	0,0060	0,5940
two-step verbal scale	<b>142</b>	117	82,39%	2,27	2,2755	0,0039	0,5538
<b>ALL</b>	<b>424</b>	339	<b>79,95%</b>	<b>2,21</b>			1,2315

Source: own research.

$$F = \frac{1}{4 - 1} \times 1,2315 = 0,41.$$

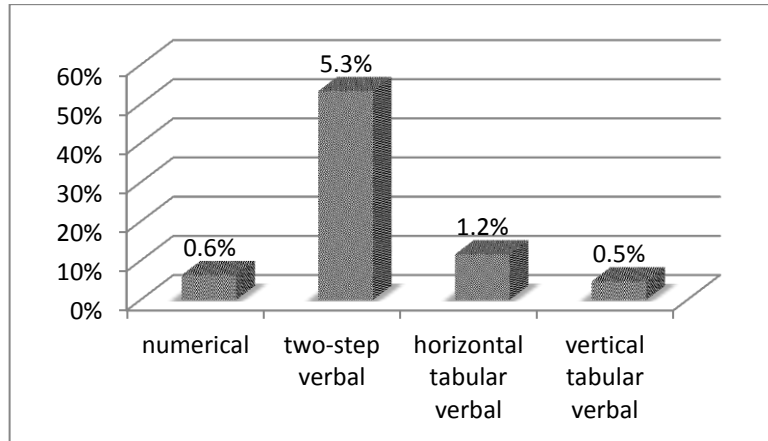
For the adopted level of significance  $\alpha=0,05$  and  $k-1$  (3) and infinity ( $\infty$ ) degrees of freedom  $F_{\alpha}$  is:  $F_{0,05} = 2,6$ . Since  $F > F_{\alpha}$ , there is no reason to reject the null hypothesis. These calculations indicate that the results of chi-square test and of F-test are the same: the graphic form of the scale does not affect the occurrence of inconsistent comparisons.

### Analysis of the correctness of completing the questionnaire depending on the form of scale

Due to a large number of questionnaires rejected due to errors in their completion, additional analysis was carried out. As mentioned above, the questionnaires were filled without individual supervision of moderator, and from 540 respondents only 424 ques-

tionnaires were included in the analysis. It has been observed that the largest number of wrong incorrectly completed questionnaires (53%) concerned numerical scale (Fig. 6). In other cases, these proportions were significantly lower.

Fig. 6. Fractions of the questionnaires by scale filled incorrectly



Source: own research.

As previously, the following hypothesis about the equality of fractions was set: the graphic form of scale has a significant impact on the appearance of incorrectly completed questionnaires. It was first verified using the chi-square test and F-test. The relevant calculations are presented in table 5.

Table 5. Empirical and expected values for chi-square test – erroneous questionnaires

Form of scale	Empirical values		N	Expected values		N
	Correct	Erroneus		Correct	Erroneus	
horizontal tabular verbal scale	105	7	<b>112</b>	87,94	24,06	<b>112</b>
numerical scale	78	89	<b>167</b>	131,13	35,87	<b>167</b>
vertical tabular verbal scale	99	13	<b>112</b>	87,94	24,06	<b>112</b>
two-step verbal scale	142	7	<b>149</b>	116,99	32,01	<b>149</b>
<b>ALL</b>	<b>424</b>	<b>116</b>	<b>540</b>	<b>424</b>	<b>116</b>	<b>540</b>

Source: own research.

As previously, the value of  $\chi^2$  was calculated, which is 147. For the adopted level of significance  $\alpha=0,05$  and for 3 degrees of freedom, the critical value  $\chi^2_{\alpha}=7,81$ . Because  $\chi^2 > \chi^2_{\alpha}$  ( $147 > 7,81$ ), the null hypothesis should be rejected in favour of the alternative one: at least one pair of fractions significantly different from each other.

It was confirmed that the graphic form of scale affects the appearance of errors in the questionnaire (which are not, however, inconsistency). Similar result was obtained using

F-test ( $F=49.52$ ,  $F_{0,05}=2.6$ ,  $F \geq F_{\alpha}$ ). Next, a post hoc test was conducted to determine those fractions, which are responsible for rejection of the null hypothesis. Six tests was performed for the following combinations of scales:

- 1) horizontal tabular verbal vs numerical
- 2) verbal tabular horizontal vs vertical tabular verbal,
- 3) horizontal tabular verbal vs. two-step verbal,
- 4) numerical vs vertical tabular verbal,
- 5) Numerical vs. two-step verbal,
- 6) vertical tabular verbal vs. two-step verbal.

Calculations are presented in Table 6.

Table 6. Results of calculations for the post-hoc test for equality of two fractions

Scale 1	Scale 2	$p$	$(1-p)$	$u$
horizontal tabular verbal scale	numerical scale	0,6559	0,3441	8,11
horizontal tabular verbal scale	vertical tabular verbal scale	0,9107	0,0893	1,41
horizontal tabular verbal scale	two-step verbal scale	0,9464	0,0536	-0,55
numerical scale	vertical tabular verbal scale	0,6344	0,3656	-7,09
numerical scale	two-step verbal scale	0,6962	0,3038	-9,38
vertical tabular verbal scale	two-step verbal scale	0,9234	0,0766	-2,08

Source: own research.

For the adopted level of significance  $\alpha=0.01$  (normal distribution):

$$u_{0,01/2}=u_{0,005} = 2,576.$$

In any case when we compare fraction of incorrectly completed questionnaire using numerical scale the null hypothesis of equal proportions must be rejected. In other cases, there is no reason to reject the null hypothesis. As the last step, chi-square tests and F-test were conducted to verify the hypothesis that other forms of scale generate equal proportions of incorrectly completed questionnaire. Using similar calculations as those described in Table 3 and 5,  $\chi^2=5$ , and the critical  $\chi^2_{\alpha}$  is 5.99 for the assumed level of significance  $\alpha=0,05$  and for 2 degrees of freedom. Hence, there is no reason to reject the null hypothesis. A similar result was obtained using the F-test ( $F=2,22$ ,  $F_{0,05}=3,0$ ,  $F < F_{\alpha}$ ). It can be concluded that numerical scale generates the largest number of incorrect questionnaires, although with respect to the consistency, there were no statistically significant differences between the graphical forms of scales.

## 6. CONCLUSIONS

The consistency ratio CR is a key indicator of consistency of judgments in the AHP/ANP methods. The value of CR should not exceed 0,10 (10%), otherwise the results should be considered erroneous (illogical), and cannot constitute a basis for a decision. For this reason, this consistency measure has long attracted attention of the researchers.

They studied both the source of this problem, as well as methods for its reduction without affecting the input data. The authors came to the obvious assumption that since the primary source of inconsistent judgments are human errors taking place at the stage of data collection, it is necessary to examine the tool by which data is collected. Generally, it is a specifically constructed questionnaire, which can be presented in different graphical forms, based on a verbal or numerical scale. The literature review revealed the existence of studies on the influence of the graphical layout of the scale on the provided responses<sup>35</sup>. However, the relationship between the consistency of judgments and the graphic form of the 9-point scale used to gather these judgments has never been investigated.

The aim of the research presented in this paper was to investigate the influence of the graphic form of the aforementioned pairwise comparisons scale on the fraction of inconsistent results (those for which  $CR > 0,10$ ). The study included 540 respondents. After rejection of erroneous questionnaires, only 424 questionnaires filled in correctly were included in further analysis. This prompted authors to perform additional analysis: whether graphical form of the scale affects the fraction of correctly completed questionnaires.

It examined four commonly used forms of scale. The AHP analysis was prepared individually for each respondent, then the fraction of responses with  $CR > 0,1$  was calculated for each of these forms. There were no significant differences between them in respect to the presence of inconsistent results ( $CR > 0,1$ ), the fraction in each case was very high with an average of approx. 80%. The performed statistical tests chi-square and F-test gave no reason to reject the hypothesis that the graphic form of the questionnaire does not affect the consistency of results within the limits recommended by Saaty. Moreover, analysis of the distribution of CR also showed no significant differences between each type of scale. It should not matter by how much CR exceeded the permissible level since each matrix with  $CR > 0,10$  must be rejected as inconsistent, and as such useless for a decision maker.

Yet the study on the relationship between CR and type of graphical presentation of scale can bring interesting conclusions regarding the choice of the scale for which the rate of CR is the lowest, even if it exceeds the level of 0,10. One would then expect that in another study, and taking into account other factors, the fraction of inconsistent judgments will be significantly lower. In the present study, there were no significant differences in this regard.

Quite different results were obtained in the analysis of the influence of the graphical form of scale on how correctly the questionnaires were completed. In this case, it was found that the largest number of wrong, incorrectly completed questionnaires (53%) was using numerical scale, which requires the respondent to enter a specific number in the blank space. Therefore, the authors recommend avoiding this type of scale in the AHP.

The main limitation of this study was their "mass scale" (each respondent filled a questionnaire on their own and in a relatively short period of time), while the AHP survey should take the form of administered questionnaire interview. Other factors that may affect the value of CR, is eg. number of degrees of the scale used (eg. 5 instead of 9). This, however, requires further experiments. It is necessary to examine a broad range of factors affecting the consistency of judgments in order to formulate the relevant recommendations allowing reduce inconsistency. The AHP/ANP research studies are usually time consuming and

---

<sup>35</sup> S.A. Webber, B. Apostolou, J.M. Hassel, *The sensitivity...*

costly (which results from the need to engage experts and the complexity of decision problems). Thus, the fact that up to 80% of the results is useless in the decision-making process because of inconsistency indicates that this area needs further research.

### Acknowledgments

The research presented in this paper was funded from the Grant of the National Science Center based on decision No. DEC-2011/01/D/HS4/04006.

### REFERENCES

- [1] Alonso J.A., Lamata M.T., *Consistency in the Analytic Hierarchy Process – A New Approach*, “International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems” 2006, Vol. 14, No. 4, pp. 445–459.
- [2] Apostolou B., Hassell J.M., *An empirical examination of the sensitivity of the analytic hierarchy process to departures from recommended consistency ratios*, “Mathematical and Computer Modelling” 1993, Vol. 17, No. 4–5, pp. 163–170.
- [3] Chu P., Liu J.K., *Note on consistency ratio*, “Mathematical and Computer Modelling” 2002, Vol. 35, No. 9–10, pp. 1077–1080.
- [4] Finan J.S., Hurley W.J., *The analytic hierarchy process: Does adjusting a pairwise comparison matrix to improve the consistency ratio help?*, “Computers & Operations Research” 1997, Vol. 24, No. 8, pp. 749–755.
- [5] Gastes D., Gaul W., *The Consistency Adjustment Problem of AHP Pairwise Comparison Matrices* [in:] A. Diamantopoulos, W. Fritz, L. Hildebrandt, eds., *Quantitative Marketing and Marketing Management*, Gabler Verlag, Wiesbaden 2012, pp. 51–62.
- [6] Ishizaka H., Labib A., *Analytic Hierarchy Process and Expert Choice: Benefits and Limitations*, “ORInsight” 2009, Vol. 22, No. 4, pp. 201–220.
- [7] Miller G.A., *The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information*, “Psychological Review” 1956, Vol. 63, No. 2, pp. 81–97.
- [8] Preston C.C., Colman A.M., *Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences*, “Acta Psychologica” 2000, Vol. 104, No. 1, pp. 1–15.
- [9] Prusak A., Stefanów P., *AHP – analityczny proces hierarchiczny. Budowa i analiza modeli decyzyjnych krok po kroku*, C.H. Beck, Warszawa 2014.
- [10] Prusak A., Stefanów P., *Badania nad właściwościami metody AHP*, Folia Oeconomica Cracoviensia, 2011, nr LII, s. 80–104.
- [11] Saaty T.L., Vargas L.G., *The Logic of Priorities*, Kluwer Nijhoff Publishing, Massachusetts 1982.
- [12] Saaty T.L., Vargas L.G., *A ratio scale metric and the compatibility of ratio scales: The possibility of arrow’s impossibility theorem*, “Applied Mathematics Letters” 1994, Vol. 7, No. 6, pp. 45–49.
- [13] Saaty T.L., *Decision Making with Dependence and Feedback. The Analytic Network Process*, RWS Publications, Pittsburgh 2001.
- [14] Saaty T.L., Ozdemir M.S., *Why the magic number seven plus or minus two*, “Mathematical and Computer Modelling” 2003, Vol. 38, No. 3–4, pp. 233–244.
- [15] Saaty T.L., Peniwati K., *Group Decision Making: Drawing out and Reconciling Differences*, RWS Publications, Pittsburgh 2007.
- [16] Saaty T.L., *Relative measurement and its generalization in decision making. Why Pairwise comparisons are central in mathematics for the measurement of intangible factors. The Analytic Hierarchy/Network Process*, „Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas” 2008, Vol. 102, No. 2, pp. 251–318.



- [17] Steczkowski J., Woźniak M., Zając K., Zeliaś A., *Statystyka matematyczna w zastosowaniach*, Akademia Ekonomiczna w Krakowie, Kraków 1996.
- [18] Stefanów P., Prusak A., *Badanie wiarygodności i skuteczności skali porównań Saaty'ego w metodzie AHP i ANP* [w:] *Przedsiębiorcze aspekty rozwoju organizacji i biznesu*, red. A. Chodyński, Oficyna Wydawnicza AFM, Kraków 2011, s. 271–298.
- [19] Targalski J., *Podejmowanie decyzji* [w:] *Organizacja i zarządzanie*, red. A. Stabryła, J. Trzcieniecki, Warszawa 1986.
- [20] Vaidya O.S., Kumar S., *Analytic hierarchy process: An overview of applications*, "European Journal of Operational Research" 2006, Vol. 169, No. 1, pp. 1–29.
- [21] Weathers D., Sharma S., Niedrich R.W., *The Impact of the Number of Scale Points, Dispositional Factors, and the Status Quo Decision Heuristic on Scale Reliability and Response Accuracy*, "Journal of Business Research" 2005, Vol. 58, pp. 1516–1524.
- [22] Webber S.A., Apostolou B., Hassel J.M., *The sensitivity of the analytic hierarchy process to alternative scale and cue presentations*, "European Journal of Operational Research" 1996, Vol. 96, No. 2, pp. 351–362.
- [23] Weijters B., Cabooter E., Schillewaert N., *The effect of rating scale format on response styles: The number of response categories and response category labels*, "International Journal of Research in Marketing" 2010, Vol. 27, pp. 236–247.

#### WPLYW FORMY 9-PUNKTOWEJ SKALI W METODZIE AHP NA SPÓJNOŚĆ OSĄDÓW

Jakość podejmowanych decyzji zależy w dużej mierze od zdolności do prawidłowego zdefiniowania i oceny problemu. W przypadku złożonych zagadnień możliwe jest i coraz częściej praktykowane stosowanie metod wspomagających decyzje, np. metod wielokryterialnych. Wymagają one jednak precyzji oraz eliminacji błędów związanych z procedurą ich wykorzystania. Celem niniejszego artykułu jest przedstawienie badań dotyczących wpływu graficznej formy dziewięciostopniowej skali porównań parami w kwestionariuszach AHP na frakcję niespójnych wyników (tzn. takich, dla których współczynnik niezgodności  $CR > 0,10$ ). Ponieważ metoda AHP jest jedną z najpopularniejszych metod stosowanych w zarządzaniu (w teorii i praktyce), w aspekcie modelowania problemów decyzyjnych, istnieje potrzeba zbadania jednego z jej najczęściej omawianych problemów – niespójności wyników. Pozwoli to na usprawnienie i podniesienie jakości podejmowania decyzji z zastosowaniem tego narzędzia. Badania miały charakter empiryczny z wykorzystaniem metody AHP. Wzięło w nich udział  $N=540$  respondentów, przy czym z powodu błędów w wypełnianiu kwestionariusza do analizy włączono jedynie 424 ankiety. Dla każdego z nich został zbudowany i zanalizowany odrębny model. Wszystkie wyniki zostały następnie wprowadzone do arkusza kalkulacyjnego i poddane analizie statystycznej. Zbadano cztery najczęściej stosowane graficzne formy skali (tj. liczbową, dwustopniową werbalną, tabelaryczną werbalną w układzie poziomym i pionowym). Testy chi-kwadrat and F-Snedecora nie wykazały istotnej różnicy między nimi, jeśli chodzi o występowanie niespójnych wyników ( $CR > 0,1$ ), co upoważnia do stwierdzenia, że forma graficzna skali nie ma wpływu na spójność udzielanych odpowiedzi. Jednakże dodatkowa analiza wykazała, że forma graficzna skali wpływa na błędy w wypełnianiu ankiet – największą liczbę błędnych ankiet (53%) zaobserwowano w przypadku skali liczbowej.

**Słowa kluczowe:** analityczny proces hierarchiczny, AHP, spójność, CR.

DOI: 10.7862/rz.2016.mmr.30

Tekst złożono w redakcji: maj 2016

Przyjęto do druku: lipiec 2016