International Journal of Information Technology & Decision Making
Vol. 20 (2021)

DOI: 10.1142/S0219622021500243

Fibonacci Series-Based Pairwise Comparison Scale for Analytic Hierarchy Process

BOĞAÇ CAN YILDIRIM* and GÜLŞAH KARAKAYA†
Department of Business Administration,
Middle East Technical University, Ankara, 06800, Turkey
*yildirim.bogac@gmail.com
†kgulsah@metu.edu.tr

MUSTAFA SINAN GÖNÜL
Newcastle Business School
Northumbria University, Newcastle, United Kingdom
sinan.gonul@northumbria.ac.uk

Published

The Analytic Hierarchy Process (AHP) is one of the most widely used quantitative tools in multi-criteria decision making problems. Despite its popularity and use due to its simple but systematic procedure, AHP has limitations especially in terms of the numerical comparison scale used in one of its core steps: pairwise comparisons. AHP is based on verbal comparisons of alternatives/criteria, which are, then, converted into quantitative scores with a one-to-one mapping between the verbal comparisons and a predetermined numerical scale. The choice of the numerical scale affects an essential characteristic of pairwise comparisons: consistency. In order to understand the intrinsic consistency propinquities, this study evaluates the most widely used numerical pairwise comparison scale (Fundamental Scale) and other numerical scales that have been proposed since the initial formulation of AHP. After identifying the limitations of known scales, a new scale based on Fibonacci series is developed considering these limitations, and further analysis is conducted through extensive simulations. The results show that the proposed scale performs well when compared to the other scales.

Keywords: Numerical comparison; AHP; comparison scale; Fibonacci; exponential; consistency

1. Introduction

Having flourished from the Operational Research discipline, Multi-Criteria Decision Making (MCDM) is concerned with providing verbal and/or computational tools to support a decision maker (DM) while he/she evaluates several alternatives with respect to multiple criteria. MCDM aims to propose a "satisfactory" solution to decision problems, as it is nearly impossible to have an optimal result in terms of all the existing criteria.

*Corresponding author.

Within this scientific field, various techniques have been proposed to systematically approach complex decision problems. Out of the techniques, Analytic Hierarchy Process (AHP) is worthy of attention because of the following three main strengths:

- representing the complex decision problem in hierarchical order that provides an overall view of the problem and helps DM to compare the items accurately[1],
- using pairwise comparison of elements rather than comparing every element at once,
- providing means to systematically check DM's consistency through pairwise comparisons.

These strengths led to its widespread use in many real-life decision making applications, such as environmental management[2], engineering[3], and location selection[4]. This, in turn, stirred up plenty of interest in academia, yielding both proponents and opponents. The latter group criticizes AHP in various aspects. Some of them[5, 6, 7, 8] focus on the scale used in pairwise comparisons, while some other studies elaborate on the weight extraction[9, 10] and consistency of evaluations[11, 12, 13]. Among these, several studies[6, 14, 15] analyzed what have already been proposed or is being widely used whereas only some studies[16] proposed novel ideas for improvement.

This study focuses on the consistency characteristics of existing numerical pairwise comparison scales. The study articulates consistency in a broader perspective rather than a narrow one that can be deduced from the findings of a few AHP applications. In order to capture the general tendency on consistency of existing numerical pairwise comparison scales, a comparative analysis and extensive experiments are conducted. To handle the limitations of known scales in terms of consistency characteristics, a new numerical pairwise comparison scale is proposed. The new scale and other existing major scales are, then, compared with the most commonly used numerical pairwise comparison scale: Saaty's Fundamental Scale. These comparisons are performed through Monte Carlo simulations for various problem settings. Rather than using few exemplary AHP applications, an extensive analysis is performed considering different problem sizes and settings for consistency. To test the performances of the scales, different performance measures, some of which are newly defined in this paper, are used. The results show that the proposed scale performs well in almost all performance measures.

The rest of this study is organized as follows: Section 2 gives an overall view of AHP procedure, numerical pairwise comparison scales, consistency concept in AHP, and the consistency-wise limitations of existing numerical pairwise comparison scales. Based on the initial analysis in Section 2, Section 3 introduces a new numerical pairwise comparison scale. In Section 4, the proposed scale and other known scales are compared with Saaty's Fundamental Scale in terms of consistency characteristics. Furthermore, implications from initial analysis are tested with detailed simulations through multiple performance measures, and the results are presented. In Section 5, initial implications, simulation results, and concluding remarks are provided. Lastly, in Section 6, limitations and further possible study areas are discussed.

2. Background and Motivation

2.1. *Procedure of AHP*

The procedure of AHP consists of five main steps:

- Definition and hierarchical representation: The main objective of the decision making problem is defined and the problem is divided into smaller manageable parts, in which, elements involved in the problem can be compared in pairs. Then, these smaller parts are grouped based on their common points and visualized on a hierarchical graph.
- Pairwise comparisons: Quantitatively incomparable elements involved in the problem are verbally compared with all other elements in its respective hierarchy level. Then, these comparisons are transcribed to numerical values using a one-to-one mapping between the verbal pairwise comparison scale and the corresponding numerical scores. These numerical scores stand for the intensity of preference of each element over the others.
- Weight derivation: Based on the numerical scores obtained in the previous steps, the "Pairwise Comparison Matrix" (PCM) is derived. Then, the PCM is processed to extract the respective weights of each criterion.
- Consistency measure: Consistency of pairwise comparisons of the DM is checked using the PCMs. If an unacceptable level of inconsistency is detected, then, the DM is advised to revise the pairwise comparisons.
- Aggregation (synthesis) of the local priorities (weights): Weights calculated using PCMs represent only the local priorities. Their contributions to the overall goal, however, still need to be determined. In the last step, previously calculated local priorities are combined to global priorities.

2.2. *Pairwise Comparison Scales in AHP*

Besides rank reversal and eigenvalue method, one of the controversial areas of AHP has been the debate over the numerical scale used in pairwise comparisons. Fundamental Scale[17, 18] is the first pairwise comparison scale used in AHP applications. The easy-to-understand logic behind the scale and its numerical simplicity have earned it a wide acceptance by AHP users. According to Ref. 3, Fundamental Scale has become the most widely used ratio scale in AHP applications.

Despite this popularity, Fundamental Scale has been criticized by researchers in many studies, and alternative scales have been proposed (Table 1). For instance, Ref. 11 points out that the scoring method's simplicity and the coarseness of the scale are inseparable in Saaty's Fundamental Scale. Although the verbal comparison part has not been a critical concern, different numerical scales have been proposed as alternatives to Fundamental Scale. Table 1 shows frequently compared scales, where, x values represent the digitized semantic part of the scales. These digitized part is used to convert verbal judgments to corresponding quantitative scale values (i.e., numerical scores) using the mathematical description of the respective scale.

Table 1 Pairwise comparison scales

Scale	Mathematical Description	Parameters (x)	Approximate Scale Values
Fundamental[17]	x	$\{1,2,\dots,9\}$	$1; 2; 3; 4;$ $5; 6; 7; 8; 9$
Power[19]	x^2	$\{1,2,\dots,9\}$	$1; 4; 9; 16; 25;$ $36; 49; 64; 81$
Root Square[19]	$\sqrt{x}$	$\{1,2,\dots,9\}$	$1; \sqrt{2}; \sqrt{3}; 2;$ $\sqrt{5}; \sqrt{6}; \sqrt{7}; \sqrt{8}; 3$
Geometric[20]	2^{x-1}	$\{1,2,\dots,9\}$	$1; 2; 4; 8; 16;$ $32; 64; 128; 256$
Inverse Linear[21]	$9/(10-x)$	$\{1,2,\dots,9\}$	$1; 1.13; 1.29; 1.5;$ $1.8; 2.25; 3; 4.5; 9$
Asymptotic[11, a]	$tanh^{-1}(\sqrt{3}(x-1)/14)$	$\{1,2,\dots,9\}$	$0; 0.12; 0.24; 0.36;$ $0.46; 0.55; 0.63; 0.7; 0.76$
Balanced[22]	$x/(1-x)$	$\{0.5, 0.55, \dots, 0.9\}$	$1; 1.22; 1.5; 1.86;$ $2.33; 3; 4; 5.67; 9$
Logarithmic[23]	$log_s(x+1)$	$\{1,2,\dots,9\}$	$1; 1.58; 2; 2.32;$ $2.58; 2.81; 3; 3.17; 3.32$

[a] It should be noted that reciprocals ($1/a_{ij}$) of the scale members (a_{ij}) are greater than 1 for Asymptotic Scale, which implies a reversed comparison score characteristic (i.e., a verbally better alternative gets a lower score). This scale has not been included in our simulations as the reverse comparison score characteristic tends to amplify the inconsistency measured in simulations.

Several other scales[8, 24, 25] have been proposed based on the identified drawbacks of the aforementioned major scales. It should be noted that our analyses and experiments only involve comparisons with the most frequently cited scales for at least two reasons: First, many of the less commonly cited scales are quite contextual and dependent on case-specific parameters (e.g. the scores of the numerical scale vary with respect to the number of alternatives and the upper limit of the corresponding scale[8]), which make them hardly suitable for our comparisons and second, for brevity and overall compatibility with other studies involving scale comparisons.

2.3. *Consistency in AHP*

By its definition, consistency in AHP requires complete transitivity in the pairwise comparison matrix. Ref. 9 mentions two kinds of transitivity:

- Ordinal Transitivity: If A is preferred to B and B is preferred to C, then A must be preferred to C.
- Cardinal Transitivity: If A is preferred to B three times and B is preferred to C twice, then A must be preferred to C six times.

A consistent matrix is defined as a matrix for which each numerical pairwise comparison score between *i* and *j* is equal to the ratio of the final weights belonging to the corresponding two elements, *i* and *j*. That is, each numerical pairwise comparison score a_{ij} in the pairwise comparison matrix is equal to

$$a_{ij} = w_i / w_j \tag{2.1}$$

where w_i and w_j are the final weights of the elements denoted on the i^{th} and the j^{th} rows in the pairwise comparison matrix, respectively. Thus, for a perfectly consistent matrix, cardinal transitivity must be satisfied.

The original application of consistency measure[17] is based on matrix perturbation theory. Ref. 17 defines consistency ratio (CR) in terms of consistency index (CI) and random index (RI)

$$CR = CI/RI \tag{2.2}$$

where

$$CI = (\lambda_{max} - n)/(n - 1) \tag{2.3}$$

In Eq. (2.2) and (2.3), λ_{max} and n are the maximum eigenvalue and the dimension of the corresponding PCM, respectively. RI is the average of CI values of randomly generated PCMs. RI is scale dependent and should be calculated for each scale separately. Ref. 17 suggests that a PCM with a maximum CR of 0.10 can be regarded as acceptably consistent.

2.4. *Limitations of Existing Pairwise Comparison Scales in Terms of Consistency*

The comparative nature of AHP needs a ratio scale for pairwise comparisons[6]. The known scales, however, have difficulties in supporting this ratio characteristic. In the commonly used ratios scale, Fundamental Scale, for instance, each adjacent major gradation in the verbal part of the scale follows an arithmetic progression. This phenomenon is named[26] as "arithmetic progression rule of the verbal part". Similarly, the numerical correspondents of the verbal scale (1, 3, 5, 7, 9) follow an arithmetic progression. Due to the reciprocity axiom of AHP, numerical correspondents below 1 (1/3, 1/5, 1/7, 1/9) form a harmonic function, which is criticized[11] as it disturbs the ratio nature of AHP. Combining all of the numerical scores yields a piecewise numerical function (Fig. 1), in which the ratio between each pair of adjacent numerical scores is different. We refer to this characteristic as "partial function characteristic" of a pairwise comparison scale. Not only Fundamental Scale but also other proposed scales (except Geometric Scale) are affected by the same feature.

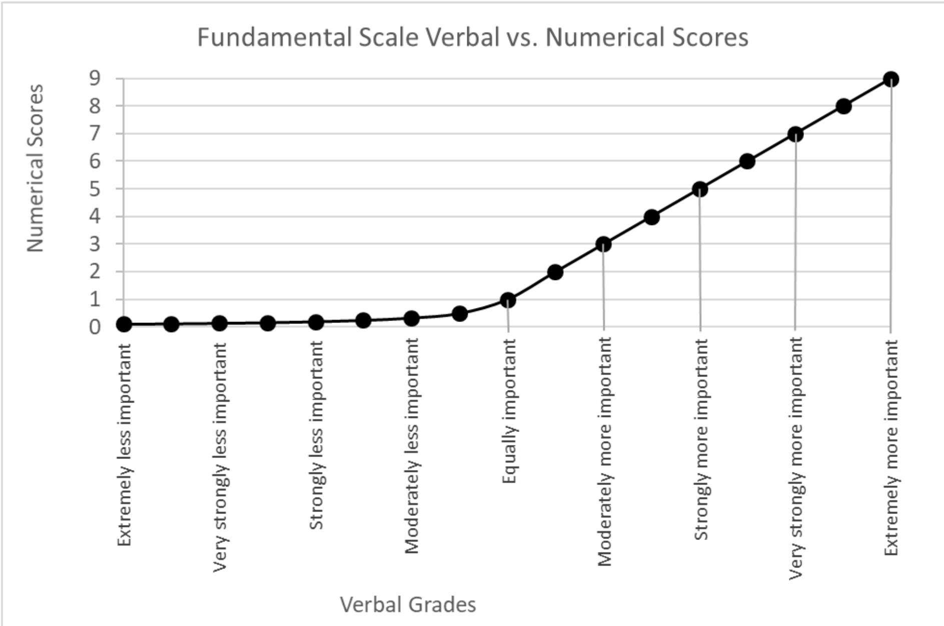

Figure 1 Piecewise numerical function for Fundamental Scale

In addition to the partial function characteristic, all of the known scales are affected by their upper and lower scale limits. Recall that consistency requires cardinal transitivity between all pairwise comparisons in an AHP application. This implies that for all a_{ij}, a_{jk}, and a_{ik}

$$a_{ik} = a_{ij} \cdot a_{jk} \tag{2.4}$$

The analysis of paired-combinations of numerical scores (a_{ij}, a_{jk}) shows that only 173 of 289 possible pairs yield an a_{ik} within the upper and lower limits of Fundamental Scale. The remaining 126 pairs, on the other hand, lie beyond the upper or lower limits. This problem of Fundamental Scale is named as the "boundary problem" and criticized by researchers[11, 26, 27]. However, Ref. 26 notes that any limited scale would suffer from this type of inconsistency tendency. Table 2 shows the extent of the boundary problem for existing pairwise comparison scales. Please refer to Appendix A Table A1 for a detailed analysis on Fundamental Scale. For the sake of consistency with the past research, this study uses numerical scores of various scales as originally suggested by their respective studies, as well as their corresponding upper and lower limits indicated in those studies.

According to Table 2, Fundamental, Logarithmic, Power, and Root Square scales are comparatively more susceptible to the boundary problem (i.e., percent multiplications within limits). While Balanced and Geometric scales seem to be better in terms of this metric, Inverse Linear scale has the highest percentage of paired combinations that fall within the upper and lower boundaries of the numerical scale.

Table 2 Boundary problem and scale discreteness analyses for different numerical scales

Scale	Number of possible combinations	Number of multiplications within limits	Percent multiplications within limits	Number of combinations corresponding to exact members within the scale	Percent combinations corresponding to exact members within the scale
Balanced	289	231	79.93%	55	19.03%
Fundamental	289	173	59.86%	85	29.41%
Geometric	289	217	75.09%	217	75.09%
Inverse Linear	289	249	86.16%	73	25.26%
Logarithmic	289	167	57.79%	61	21.11%
Power	289	173	59.86%	85	29.41%
Root Square	289	173	59.86%	85	29.41%

Another critical problem regarding the cardinal transitivity requirement is that the multiplication of pairs does not always result in a numerical value on the scale. Suppose a_{ij} = 8 and a_{jk} = 1/3. This combination is transitive if and only if a_{ik} = 8/3. Note that the suggested value of a_{ik} is not a member within Fundamental Scale. Obviously, such combinations may never satisfy cardinal transitivity, giving the scale an inherent propensity for inconsistency. This limitation is mainly attributed to scale discreteness.

Ref. 26 shows that 44 of 81 possible paired-combinations are not members within Fundamental Scale. However, this analysis only examines the combinations of numerical scores between 1-9 with the numerical scores between 1/9-1. Extending the analysis to the entire numerical value range of Fundamental Scale shows that 204 of 289 possible paired-combinations result in a numerical value that is not a member within the scale. Except for Geometric Scale, all scales appear to be highly susceptible to this inconsistency propensity due to scale discreteness. Columns 5 and 6 of Table 2 show the extent of this characteristic for existing pairwise comparison scales.

According to Ref. 26, an AHP scale can uphold the transitivity if

- its verbal part satisfies the arithmetic progression rule, and
- its numerical part satisfies the geometric progression rule.

Whether the verbal pairwise comparison scale satisfies an arithmetic progression or not is out of the scope of this study. Therefore, the discussion boils down to the geometric characteristics of numerical pairwise comparison scales.

Geometric progression of the numerical pairwise comparison scale implies that for each element a_n in a series A

$$a_n/a_{n+1} = a_{n+1} / a_{n+2} \tag{2.5}$$

Based on this requirement, a generalized geometric series can be formulated as

$$a_n = a_1 \cdot b^{(n-1)} \tag{2.6}$$

where a_n is the n^{th} element of the series and b is the common ratio between the successive elements. Note that Geometric Scale[20] ($a_n = a_1 \cdot 2^{(n-1)}$) satisfies the geometric progression rule. Additionally, exponential series ($a_n = a_1 \cdot e^{(n-1)}$) also satisfies the geometric progression. Although possible exponential series have been mentioned by Ref. 20, to our knowledge none has been adapted for AHP applications.

3. The Proposed Method

3.1. *The Idea*

An exponential pairwise comparison matrix is first proposed[20] with the following generalized formula

$$a_{ij} = e^{\lambda\delta_{ij}} \tag{3.1}$$

where a_{ij} is the numerical score from the comparison between elements *i* and *j*, δ_{ij} is an integer designating the gradation chosen by the DM to estimate the preference ratio between elements *i* and *j*, and λ is the scale constant. Each verbal comparison of a DM is converted to a numerical grade of δ_{ij} by one-to-one mapping between the verbal comparisons and the numerical scale. Given the scale constant λ each verbal comparison has a numerical score a_{ij}.

Ref. 20 notes that there is no unique value for scale constant λ and suggests that $\lambda = 1$ or $\lambda = 2$ would be appropriate choices. This intuitive selection roots neither from a numerical formulation nor an experimental evidence that suggests the abovementioned values. The intuitiveness in this suggestion might be one of the main reasons for exponential scale being unable to become as popular as the other scales.

3.2. *Exponential Scale Based on Fibonacci Series*

The Fibonacci series consists of integers where a number in the series is equal to the sum of the previous two numbers. This rule is mathematically formulated as

$$F_n = F_{n-1} + F_{n-2} \tag{3.2}$$

where the first two numbers in the series are initially defined as $F_1 = 1$ and $F_2 = 1$. Then, the elements of the series are as follows

$$1 \quad 1 \quad 2 \quad 3 \quad 5 \quad 8 \quad 13 \quad 21 \quad 34 \quad 55 \quad 89 \ldots$$

If we disregard the initially defined first element, the elements in Fibonacci series have an exponential characteristic due to the nature of its formulation (Eq. 3.2). Reciprocity condition of AHP implies that for all *i* and *j*

$$a_{ij} = 1/a_{ji} \tag{3.3}$$

Parametrizing the mathematical function (Eq. 3.1) to satisfy this condition is possible by setting δ_{ij} values to integers around 0. When $\delta_{ij} = 0$, the function yields a numerical score of 1, which is the numerical correspondent of "equal importance" in a verbal pairwise comparison. In order to represent the function in an easily understandable format, the one-to-one mapping between the verbal scale and numerical grades is shown in Table 3.

Table 3 One-to-one mapping between the verbal scale and numerical grades in exponential function

Verbal Grades	Numerical Grades (δ_{ij})
Equally important	0
Slightly more (less) important	2 (-2)
Strongly more (less) important	4 (-4)
Very strongly more (less) important	6 (-6)
Extremely more (less) important	8 (-8)
Compromises	±1, ±3, ±5, ±7

Based on the given mapping, the elements of Fibonacci series fit very well on an exponential function as plotted in Fig. 2.

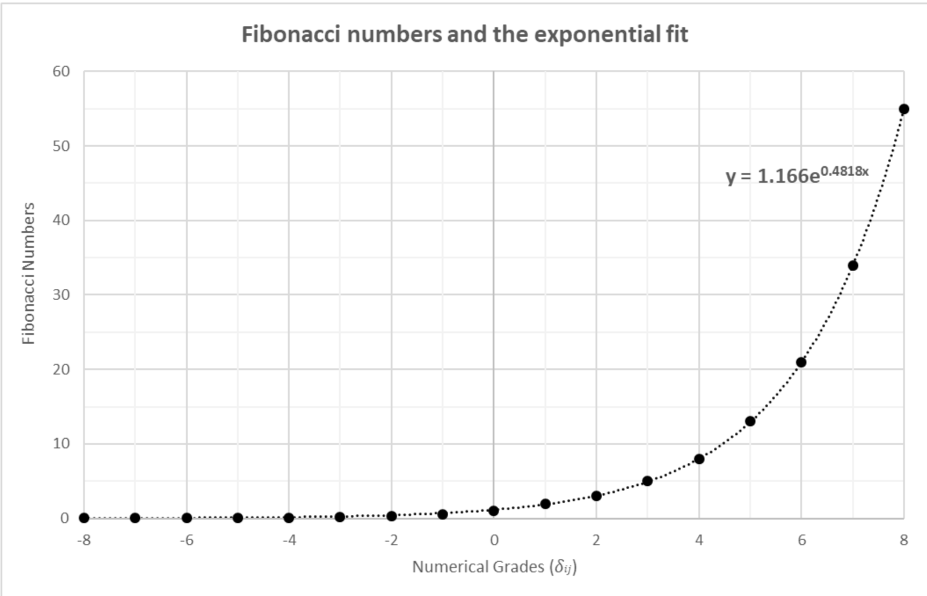

Figure 2 The exponential function fit to the elements of Fibonacci series and their reciprocals

Fig. 2 shows that the function formulated in Eq. 3.4 fits the points with an *R-square* value of "1.00". However, the coefficient in the function results in an a_{ij} value of "1.166" even for equal importance situation, where $\delta_{ij} = 0$. The function is divided by the coefficient itself to achieve the final equation as

$$a_{ij} = 1.166e^{0.4818\delta_{ij}} / 1.166 = e^{0.4818\delta_{ij}} \tag{3.4}$$

The abovementioned formula yields a unique value for each verbal comparison and corresponding numerical grade. This mapping is shown in Table 4.

Table 4 One-to-one mapping of numerical values for each verbal scale gradation and its corresponding numerical grade

Verbal Scale	Numerical Grade (δ_{ij})	Approximate Numerical Value (a_{ij})
Extremely less important	-8	0.021
	-7	0.034
Very strongly less important	-6	0.056
	-5	0.090
Strongly less important	-4	0.146
	-3	0.236
Slightly less important	-2	0.382
	-1	0.618
Equally important	0	1.000
	1	1.619
Slightly more important	2	2.621
	3	4.244
Strongly more important	4	6.870
	5	11.123
Very strongly more important	6	18.008
	7	29.154
Extremely more important	8	47.200

With the mapping between the verbal scale and the corresponding numerical values for pairwise comparisons, the function becomes a numerical pairwise comparison scale. This scale is referred to as “Fibonacci Series-Based Exponential Scale” (FSBES) in the rest of the study.

The exponential characteristic of the new scale brings several advantages. The first advantage is that the function is not a piecewise function and the ratio between each adjacent pair of elements is constant, as shown in Eq. 3.5. AHP, by its nature, makes use of ratio scales for numerical score assessments. FSBES may be more compatible to AHP as a ratio scale than those with partial functions and variable ratios.

$$e^{0.4818 \cdot (n+1)} / e^{0.4818 \cdot n} = e^{0.4818} \tag{3.5}$$

Secondly, FSBES is extendable without deforming the constant ratio property, mainly because it is defined as a continuous function with geometric progression. However, one should still keep in mind that the boundary problem exists for all pairwise comparison scales with upper and lower bounds. Table 5 exhibits the extent of the boundary problem and inherent inconsistency due to scale discreteness for FSBES as well as Fundamental

Scale as a benchmark. FSBES performs well in terms of both boundary problem and scale discreteness. The reader is referred to Table 2 for the performances of other scales.

Table 5 Possible paired-combination results within limits of the scale and/or defined by the scale (Fundamental Scale vs. FSBES)

Scale	Number of possible combinations	Number of multiplications within limits	Percent multiplications within limits	Number of combinations corresponding to exact members within the scale	Percent combinations corresponding to exact members within the scale
Fundamental	289	173	59.86%	85	29.41%
FSBES	289	217	75.09%	217	75.09%

Another advantage is that the multiplication of each paired-combination (a_{ij}, a_{jk}) is an element of the main function, as shown in Eq. 3.6.

$$a_{ik} = a_{ij} \cdot a_{jk} = e^{0.4818(\delta_{ij}+\delta_{jk})} \tag{3.6}$$

Multiplication results being elements of the main function eliminate the inherent inconsistency tendency due to scale discreteness. Especially in large PCMs, the likelihood of satisfying cardinal transitivity decreases due to scale discreteness, even if ordinal transitivity is completely satisfied. In this regard, a numerical scale with a constant ratio between successive elements may be more likely to generate a higher consistency than that of variable ratios.

It is also worth mentioning that Geometric Scale has a relatively high performance in inherent inconsistency tendency, just as FSBES. This indication can be attributed to the continuous functions and constant ratios that reside under the bases of these scales.

4. Computational Experiments

AHP has been extensively discussed since its initial formulation. However, most of these discussions were either confined within theoretical borders or based only on a practical application of AHP to a selected decision problem. The conclusions on the general consistency characteristics drawn from such studies, thus, remain highly limited and narrow in scope.

On this regard, this study utilizes an inherently more generalizable, empirical analysis based on simulations of randomly generated scenarios. The analyses were carried out by generating large numbers of random PCMs to compare consistency tendencies of pairwise comparison scales with that of Fundamental Scale. Fig. 3 shows the flow of the algorithm starting from the generation of random PCMs to performance measurements and generalization phases. It is also worth mentioning that this approach can be used in various combinations of local weight calculation methods and error approximation methods.

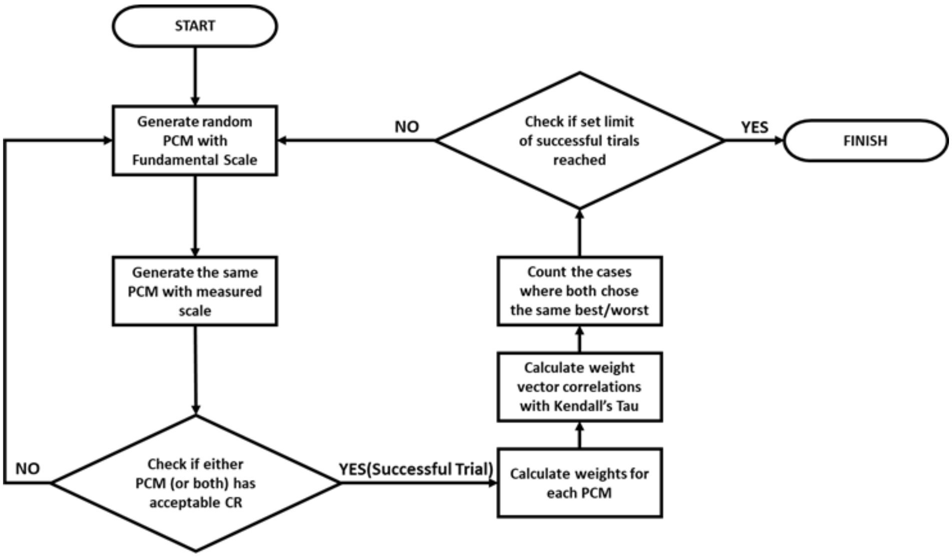

Figure 3 Basic algorithm representing the process of performance measurement

Each scale is tested through this algorithm for various matrix dimensions (3x3 to 7x7) and different CR acceptance limits (0.15, 0.10, 0.05). Each scale is compared against Fundamental Scale, as it is the most widely used pairwise comparison scale[28]. Then, the results were compared and evaluated via multiple performance metrics:

- CR Comparison: These metrics represent the percentages of PCMs, where the measured scale or Fundamental Scale generated a more consistent PCM, or the cases with equal CR values.
  - Measured Scale Lower CR: The percentage of successful trials for which the use of measured scale resulted in a lower CR value
  - Fundamental Scale Lower CR: The percentage of successful trials for which the use of Fundamental Scale resulted in a lower CR value
  - Equal CR: The percentage of successful trials for which both measured scale and Fundamental Scale resulted in the same CR value
- Inconsistency Percentages: The simulation algorithm was formulated such that even if one of the compared scales generates an acceptably consistent PCM, then the trial is successful. This limit is used in order to avoid data disturbance resulting from unacceptably inconsistent matrices for both scales. Once the trial is successful, then it is assigned to one of the below-mentioned categories in order to give the reader a more detailed understanding of the consistency tendencies of different scales.
  - Fundamental Scale Inconsistent: The percentage of successful trials where Fundamental Scale generated intolerably inconsistent PCMs
  - Measured Scale Inconsistent: The percentage of successful trials where measured scale generated intolerably inconsistent PCMs
  - Both Scales Consistent: The percentage of successful trials where both scales generated consistent PCMs

- Choosing the Same Best/Worst: This group of metrics represents the respective percentages of the successful trials, in which, both of the compared scales point out the same alternative as the best/worst. These percentages indicate how close the results would be if the aim was only to select one of the alternatives as the best/worst.
  - Same Best Chosen: The percentage of successful trials where both scales selected the same item as the best choice
  - Same Worst Chosen: The percentage of successful trials where both scales selected the same item as the worst choice
- Overall Ranking Similarity: In case the DM is not only looking for the single best option but aims to divide the resources based on the weights, all of the individual weights become important. Therefore, it is also valuable to see how much the weight vectors differ based on the selected scale. The metrics in this group helps the reader see how close the entire set of resultant weights to those generated by Fundamental Scale.
  - Kendall's Tau Mean: The average ranking correlation of weight vectors obtained from the same PCM generated by different scales (Fundamental Scale and measured scale), in terms of Kendall's Tau correlation coefficient
  - Kendall's Tau Standard Deviation: Standard deviation of Kendall's Tau values for all successful trials

Abovementioned performance metrics are reported as the average results of 1000 successful trials for each "measured scale – matrix dimension – CR limit" combination. For each scale, RI values are calculated by simulations and respective RIs are used for CR calculations. Table 6 shows the performance results of FSBES in 3x3 PCMs against Fundamental Scale. The first column represents the dimension of PCMs. For each dimension, three different acceptable CR limits are used in order to assess the CR limit sensitivity of the results. The results of each CR limit have been tabulated on the corresponding row. Columns 3 and 4 show the percentage of the PCMs where the measured scale and Fundamental Scale resulted in a lower CR value, respectively, and column 5 shows the percentage of the PCMs where CRs were equal for both scales. Columns 6 and 7 show the percentages where one scale was consistent while the other one was inconsistent. Column 8 shows the percentage successful trials, where both scales generated acceptably consistent PCMs. According to Table 6, FSBES shows better consistency characteristics as it resulted in lower CR values. Moreover, it can also be deduced that FSBES tends to create more consistent PCMs even when the PCM formed by Fundamental Scale is inconsistent. This trend is observed more evidently with increasing PCM dimensions.

Although FSBES results in lower CR values in general, it should still be checked if it actually creates irrelevant PCMs. Therefore, not only the CRs but also the final rankings pointed by both scales have been checked. Columns 9 and 10 show percentages of PCMs that pointed the same best and worst elements for both numerical comparison scales. Columns 11 and 12 show the comparison results of entire rankings in all successful PCM pairs. For these comparisons, Kendall's Tau rank correlation coefficients[29] have been calculated. Kendall's Tau rank correlation coefficient (τ) takes values between -1 (complete disagreement between two rankings) and +1 (complete agreement between two rankings). Simulation results indicate that the average Kendall's Tau values are very close to +1 for FSBES, which shows that the rankings with FSBES do not differ significantly from those with Fundamental Scale. Although ranking preservation has been observed to decrease with increasing matrix dimension, the percentages still remain well above 85% showing a high correspondence rate with Fundamental Scale results. Detailed results regarding all scales can be seen in Appendix B.

Table 6 Performance results of FSBES in comparison to Fundamental Scale (3x3 matrices)

n	CR Limit	Measured Scale Lower CR	Fund. Scale Lower CR	Equal CR	Fund. Scale Incons.	Measured Scale Incons.	Both Scales Consistent	Same Best Chosen	Same Worst Chosen	Kendall's Tau Mean	Tau Standard Deviation
3x3	0.15	67.00%	32.90%	0.10%	25.00%	15.40%	59.60%	96.40%	97.30%	0.9555	0.1643
	0.10	68.90%	31.10%	0.00%	41.00%	11.90%	47.10%	97.40%	97.00%	0.9608	0.1540
	0.05	65.50%	34.40%	0.10%	54.70%	17.40%	27.90%	100.00%	99.80%	0.9961	0.0367

Fig. 4 shows the CR performances of different scales for each matrix dimension from 3x3 to 7x7. Data points represent the percentage of trials for which the PCM generated by the measured scale resulted in a lower CR value than that of Fundamental Scale. Apparently Logarithmic and Root Square scales have very poor performance in terms of CR when compared to Fundamental Scale. The other scales tend to have better CR characteristic with increasing matrix dimension. While Power Scale appears to be the top performer in this measure, Geometric Scale and FSBES reach Power Scale's performance. Balanced and Inverse Linear scales seem to be somewhat slow to catch up with the others. For the sake of simplicity, only the performances for CR limit of 0.10 are visualized in Fig.4.

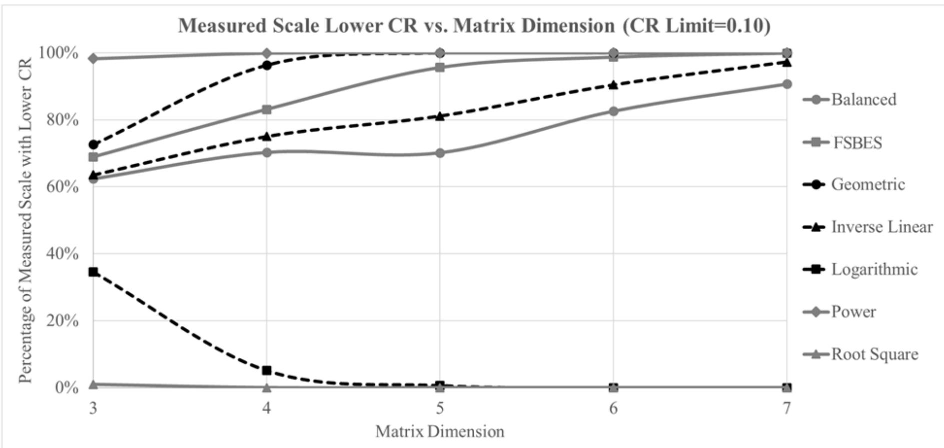

Figure 4 Performances of different scales in terms of generating lower CR PCMs (CR Limit=0.10)

Averaging the directly measurable five performance criteria of each scale over all matrix dimensions from 3x3 to 7x7 results in an overall performance of the scales for the corresponding CR limit. Fig. 5 shows the map chart that illustrates the results for each scale. Apparently, all measured scales are somewhat aligned with Fundamental Scale in terms of weight vector related measures (Kendall's Tau Mean, Same Best/Worst Chosen). In terms of consistency, on the other hand, FSBES, Geometric, and Power scales appear to be the ones with the highest performance.

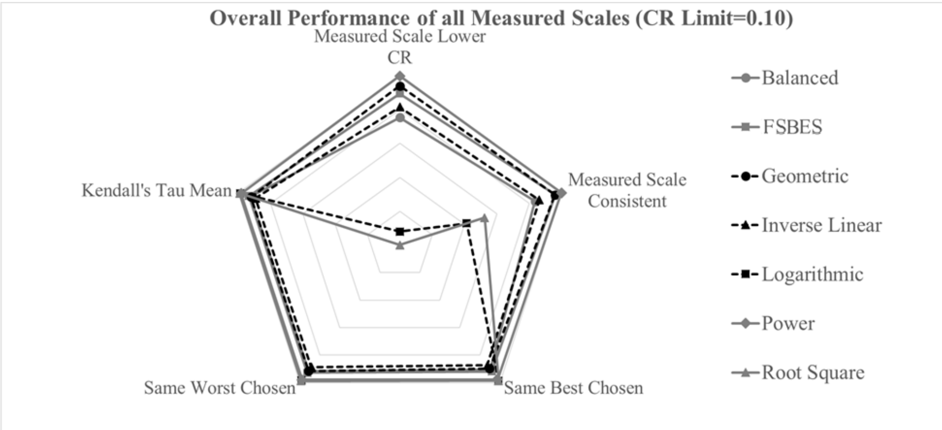

Figure 5 Comparison of scales based on averaged performance scores (CR Limit=0.10)

Although Fig. 5 demonstrates the results for CR=0.10 case, the performances are also similar for CR limits of 0.05 and 0.15.

5. Discussion and Concluding Remarks

Despite being the most widely used pairwise comparison scale in AHP applications, our study indicates that Fundamental Scale tends to introduce large inconsistencies into PCMs. First, although numerical comparisons in AHP are based on ratio comparisons in theory, the ratios between successive elements vary in Fundamental Scale, which can be considered as impedimentary.

Fundamental Scale also tends to bring along two types of different inconsistency characteristics to PCMs. The first inconsistency introduced by Fundamental Scale is related to the boundary problem. This issue has previously been addressed[26] in the literature, however, the analysis was somewhat limited in terms of possible scenarios. This study extends the boundary problem analysis in a more generalized way with further improvement possibilities. Our analysis also indicates that all of the evaluated scales are eventually vulnerable to the boundary problem, while some scales (including Fundamental Scale) are significantly more susceptible than others.

The second and more critical inconsistency characteristic roots from scale discreteness, which means that the multiplication of some of the possible numerical score pairs are not actually members within the scales themselves. This characteristic is particularly important as cardinal transitivity is ultimately necessary for perfect consistency in an ideal case. The multiplications of pairs not being members within the scale, on the other hand, contradicts with cardinal transitivity. It is expected that the scales with geometric progression (or similar to geometric progression) eliminate this inconsistency propensity.

Geometric and exponential type series satisfy the geometric progression rule, thus, they may bring an advantage to decrease the unwittingly introduced inconsistencies based on the nature of the numerical pairwise comparison scale. An earlier proposal for an exponential type scale did not attract much attention. The very first idea of an exponential type scale has been formulated on an intuitive basis, which may have played a role in its unpopularity.

One of the most famous series that largely satisfy the exponential progression rule is the Fibonacci series. With slight adjustments, Fibonacci series can be formulated on exponential basis and used as a numerical pairwise comparison scale in AHP. This new scale is called FSBES.

When compared with Fundamental Scale, proposed FSBES has three main advantages:

- Ratios between successive elements of the scale are constant.
- Possible paired combination multiplications are more likely to be within the upper and lower limits.
- Possible paired combination multiplications are more likely to correspond to the existing values within the scale.

Considering these advantages, FSBES is expected to result in generating more consistent PCMs, when the same verbal comparisons are used as inputs.

In line with the preliminary expectations, our simulations show that FSBES has superior performance in terms of consistency when compared to Fundamental Scale. Another notable finding is that FSBES has better consistency properties without deviating significantly from Fundamental Scale's results. We note that despite the abovementioned issues, Fundamental Scale is still the most commonly used numerical pairwise comparison scale, thus an important benchmark. FSBES provides very similar results to those of the common and proven practice, i.e., Fundamental Scale. Additionally, more consistent PCMs ultimately mean that FSBES relieves the burden of revising PCMs for multiple times, thus improving AHP's practical applicability.

Our results also indicate that FSBES's comparative consistency performance becomes much better with increasing PCM dimensions. The change in the CR limits for the same matrix dimension, on the other hand, does not have a significant effect on the consistency performance indicators. Based on ranking comparisons, it is observed that the increasing PCM dimensions slightly decreased the ranking correlations of the PCMs generated by Fundamental Scale and FSBES. Keeping the matrix dimension constant, decreasing the CR limit led to higher correlations between the PCMs generated by Fundamental Scale and FSBES.

Similar to FSBES, Geometric and Power scales have good consistency performance scores. Yet, it should be noted that the most important problem for the Geometric Scale[20] is that the upper limit of 256 severely violates[26] the homogeneity axiom of AHP. Although not as much as for Geometric Scale, the upper limit of 81 in Power Scale also violates the same axiom. FSBES, on the other hand, is more acceptable with respect to the homogeneity axiom, when compared to the other scales with geometric progression. Therefore, it may be more preferable to use FSBES when computational limitation of the human mind is considered.

6. Limitations and Future Research

Our study is limited to the comparison of the most frequently discussed numerical scales with Saaty's Fundamental Scale. Therefore, the results do not necessarily indicate the performances of all of the scales against one another. Further research may be focused on more detailed analysis on paired comparisons of the scales to investigate their overall performances. Additionally, further research can be extended to include more numerical pairwise comparison scales including the recently developed and less frequently used ones.
In our simulations, the weight vectors of PCMs are calculated through Eigenvalue Method (EVM). According to Ref. 30, Mean of Normalized Values is frequently used to approximate the EVM in many practical applications. Similarly, Row Geometric Mean Method[31] is seen as a strong alternative to EVM. Future research may examine the effects of the use of different weight extraction methods on scale-wise consistency performances.
Similarly, consistency performances are calculated through EVM. For these calculations, RI values for each scale have been generated through Monte Carlo simulations in line with literature[32]. Future research may make use of alternative consistency measure techniques, such as Geometric Consistency Index (GCI).
Another important point to consider is the behavior of a DM against variation of numerical scale. Does a DM change the verbal judgement if the numerical scale changes? A rational DM, who knows what numerical value corresponds to his/her verbal judgements may not use the same verbal evaluations for different scales. In such a case, the DM may adjust the verbal evaluations so that the numerical judgements represent his/her actual idea of the ratio weights. Thus, we believe that the use of the same verbal scale should be carefully considered in numerical scale comparisons.
The above-mentioned limitation applies to our study, as well. In our simulations, PCMs are generated based on Fundamental Scale, and then, numerical scores in the PCMs are converted to the corresponding numbers in other scales. Effectively, this means the use of same verbal scale for both Fundamental and the measured scale. Further research may investigate the DM's perception of the verbal scale when different numerical scales are used.
Another limitation worth mentioning is that simulation durations significantly increased with increasing PCM dimensions and decreasing CR limits. For the sake of time and simplicity, our simulations were run up to 7x7 PCMs. Further study may focus on

generating a larger mapping to see if the patterns found in our study would change for larger PCMs. In order to preserve the randomness of the PCM generation, we did not introduce any bias in the selection of scores. However, several existing methods introduce bias on the random selection based on the choice of the first comparison score. Simulation durations may be decreased by introducing bias towards the first selection, yet, it may disturb the overall randomness of PCMs, and thus, the results may differ from ours'.

It is also worth mentioning that the base parameter "$e^{0.4818}$" in Eq. 3.4 has an approximate value of 1.619, which is very close to the so-called "Golden Ratio" (~1.618). Future research can also focus on the analysis of the Golden Ratio in strength of preference between alternatives.

Finally, the current study adopted an inductive approach where major scales were first analyzed based on initial findings with regards to their consistency characteristics. Then, these findings have been thoroughly tested through wide-encompassing and extensive simulations. This, in our opinion, brings along more universal and agreeable results compared to generalizing from a single case study based on a unique AHP application. Arguably, even in this form, the results remain fairly limited in scope and cannot be used to derive a theory nor improve an existing one. Thus, future studies should embrace a deductive approach where they start from an existing theory and employ a top-down approach to design specific methods and techniques to improve the theory and practice of AHP.

References

1. A. El Hefnawy and A. S. Mohammed, Review of Different Methods for Deriving Weights in the Analytic Hierarchy Process, *International Journal of the Analytic Hierarchy Process*, **6**(1) (2014) 95-123.
2. L. Z. Stirn and P. Grošelj, Multiple Criteria Methods with Focus on Analytic Hierarchy Process and Group Decision Making, *Croatian Operational Research Review (CRORR)*, **1** (2010) 2-11.
3. S. J. Chen and L. Lin, Decomposition of Interdependent Task Group for Concurrent Engineering, *Computers & Industrial Engineering*, **44**(3) (2003) 435-459.
4. Y. Takamura and K. Tone, A Comparative Site Evaluation Study for Relocating Japanese Government Agencies in Tokyo, *Socio-Economic Planning Sciences*, **27**(2) (2003) 85-102.
5. Y. Dong, Y. Xu, H. Li and M. Dai, A Comparative Study of the Numerical Scales and the Prioritization Methods in AHP, *European Journal of Operational Research* **186**(1) (2008) 229-242.
6. J. Franek and A. Kresta, Judgment Scales and Consistency Measure in AHP, *Procedia Economics and Finance* **12**(1) (2014) 164-173.
7. C. Garuti, Reflections on scales from measurements, not measurements from scales, *International Journal of the Analytic Hierarchy Process* **9**(3) (2017) 349–361.
8. K. D. Goepel, Comparison of Judgment Scales of the Analytical Hierarchy Process - A New Approach, *International Journal of Information Technology & Decision Making* **18**(2) (2019) 445-463.
9. T. L. Saaty and G. Hu, Ranking by the Eigenvector Versus Other Methods in the Analytic Hierarchy Process, *Applied Mathematical Letters*, **11**(4) (1998) 121-125.
10. T. K. Dijkstra, On the Extraction of Weights from Pairwise Comparison Matrices, *Central European Journal of Operations Research*, **21**(1) (2013) 103-123.
11. F. J. Dodd, H. A. Donegan and T. B. M. McMaster, A New Approach to AHP Decision-Making, *Journal of the Royal Statistical Society*, Series D (The Statistician), **41**(3) (1992) 295-302.
12. A. Davoodi, On Inconsistency of a Pairwise Comparison Matrix, *International Journal of Industrial Mathematics*, **1**(4) (2009) 343-350.
13. J.B. Kim, Y.G. Cho, Y.B. Kim and K.T. Cho, New Criteria for the Consistency in Reasonable Pairwise Comparison Matrices, *Journal of the Korean Institute of Industrial Engineers*, **36**(1) (2010) 1-6.
14. A. Ishizaka and M. Lusti, How to Derive Priorities in AHP: A Comparative Study, *Central European Journal of Operations Research*, **14**(4) (2006) 387-400.
15. G. Kou, D. Ergu, C. Lin and Y. Chen, Pairwise Comparison Matrix in Multiple Criteria Decision Making, *Technological and Economic Development of Economy*, **22**(5) (2016) 738-765.
16. C. Lin, G. Koua and D. Ergu, A Heuristic Approach for Deriving the Priority Vector in AHP, *Applied Mathematical Modelling*, **37**(8) (2013) 5828-5836.
17. T. L. Saaty, *The Analytic Hierarchy Process*, (McGraw Hill International, New York, 1980).
18. T. L. Saaty, *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy*, (RWS Publications, Pittsburgh, 1996).
19. P. Harker and L. Vargas, The theory of ratio scale estimation: Saaty's analytic hierarchy process, *Management Science* **33**(11) (1987) 1383-1403.
20. F. A. Lootsma, Conflict Resolution via Pairwise Comparison of Concessions, *European Journal of Operational Research*, **40**(1) (1989) 109-116.
21. D. Ma and X. Zheng, 9/9–9/1 Scale method of AHP, *2nd Int. Symp. AHP*, Vol. 1 (Pittsburgh, 1991), pp. 197-202.
22. A. Salo and R. Hämäläinen, On the measurement of preferences in the analytic hierarchy process, *J. Multi-Criteria Decision Analysis* **6** (1997) 309-319.

23. A. Ishizaka, D. Balkenborg and T. Kaplan, Influence of aggregation and measurement scale on ranking a compromise alternative in AHP, *Journal of the Operational Research Society* **62** (2010) 700-710.
24. M. Elliott, Selecting numerical scales for pairwise comparisons, *Reliability Engineering and System Safety*, **95** (2010) 750-763.
25. W.W. Koczkodaj, Pairwise comparisons rating scale paradox, in Transactions on Computational Collective Intelligence XXII, eds., N. T. Nguyen and R. Kowalczyk, *Lecture Notes in Computer Science*, Vol. 9655 (Springer, Berlin, Heidelberg, 2016), 1-9.
26. P. Ji and R. Jiang, Scale Transitivity in the AHP, *The Journal of the Operational Research Society*, **54**(8) (2003) 896-905.
27. S. French, *Decision Theory: An Introduction to the Mathematics of Rationality*, (Ellis Horwood Ltd., West Sussex, 1988).
28. Z. Zhang, X. Liu and S. Yang, A Note on the 1-9 Scale and Index Scale In AHP, *Communications in Computer and Information Science*, **35**(1) (2009) 630-634.
29. W. W. Daniel, *Applied nonparametric statistics, The Duxbury advanced series in statistics and decision sciences*, (PWS-Kent, Boston, 1990).
30. E. Mu and M. Pereyra-Rojas, *Practical Decision Making: An Introduction to the Analytic Hierarchy Process (AHP) Using Super Decisions V2*, (Springer Online Documents: Springer International Publishing, pp. 7-22, 2017).
31. E. Triantaphyllou and S. H. Mann, Using The Analytic Hierarchy Process For Decision Making In Engineering Applications: Some Challenges, *International Journal of Industrial Engineering: Applications and Practice*, **2**(1) (1995) 35-44.
32. J. A. Alonso and M. T. Lamata, Consistency in the Analytic Hierarchy Process: A New Approach, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **14**(4) (2006) 445-459.

Appendix A. Consistency Characteristics Analysis

Table A 1 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in Fundamental Scale

	9.000	**8.000**	**7.000**	**6.000**	**5.000**	**4.000**	**3.000**	**2.000**	**1.000**	**0.500**	**0.333**	**0.250**	**0.200**	**0.167**	**0.143**	**0.125**	**0.111**
9.000	81.000	72.000	63.000	54.000	45.000	36.000	27.000	18.000	9.000	4.500	3.000	2.250	1.800	1.500	1.286	1.125	1.000
8.000	72.000	64.000	56.000	48.000	40.000	32.000	24.000	16.000	8.000	4.000	2.667	2.000	1.600	1.333	1.143	1.000	0.889
7.000	63.000	56.000	49.000	42.000	35.000	28.000	21.000	14.000	7.000	3.500	2.333	1.750	1.400	1.167	1.000	0.875	0.778
6.000	54.000	48.000	42.000	36.000	30.000	24.000	18.000	12.000	6.000	3.000	2.000	1.500	1.200	1.000	0.857	0.750	0.667
5.000	45.000	40.000	35.000	30.000	25.000	20.000	15.000	10.000	5.000	2.500	1.667	1.250	1.000	0.833	0.714	0.625	0.556
4.000	36.000	32.000	28.000	24.000	20.000	16.000	12.000	8.000	4.000	2.000	1.333	1.000	0.800	0.667	0.571	0.500	0.444
3.000	27.000	24.000	21.000	18.000	15.000	12.000	9.000	6.000	3.000	1.500	1.000	0.750	0.600	0.500	0.429	0.375	0.333
2.000	18.000	16.000	14.000	12.000	10.000	8.000	6.000	4.000	2.000	1.000	0.667	0.500	0.400	0.333	0.286	0.250	0.222
1.000	9.000	8.000	7.000	6.000	5.000	4.000	3.000	2.000	1.000	0.500	0.333	0.250	0.200	0.167	0.143	0.125	0.111
0.500	4.500	4.000	3.500	3.000	2.500	2.000	1.500	1.000	0.500	0.250	0.167	0.125	0.100	0.083	0.071	0.063	0.056
0.333	3.000	2.667	2.333	2.000	1.667	1.333	1.000	0.667	0.333	0.167	0.111	0.083	0.067	0.056	0.048	0.042	0.037
0.250	2.250	2.000	1.750	1.500	1.250	1.000	0.750	0.500	0.250	0.125	0.083	0.063	0.050	0.042	0.036	0.031	0.028
0.200	1.800	1.600	1.400	1.200	1.000	0.800	0.600	0.400	0.200	0.100	0.067	0.050	0.040	0.033	0.029	0.025	0.022
0.167	1.500	1.333	1.167	1.000	0.833	0.667	0.500	0.333	0.167	0.083	0.056	0.042	0.033	0.028	0.024	0.021	0.019
0.143	1.286	1.143	1.000	0.857	0.714	0.571	0.429	0.286	0.143	0.071	0.048	0.036	0.029	0.024	0.020	0.018	0.016
0.125	1.125	1.000	0.875	0.750	0.625	0.500	0.375	0.250	0.125	0.063	0.042	0.031	0.025	0.021	0.018	0.016	0.014
0.111	1.000	0.889	0.778	0.667	0.556	0.444	0.333	0.222	0.111	0.056	0.037	0.028	0.022	0.019	0.016	0.014	0.012
Color Codes	Within the limits and defined by the scale							Within the limits of the scale					Outside the limits of the scale				

Table A 2 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in FSBES

	47.200	**29.154**	**18.008**	**11.123**	**6.870**	**4.244**	**2.621**	**1.619**	**1.000**	**0.618**	**0.382**	**0.236**	**0.146**	**0.090**	**0.056**	**0.034**	**0.021**
47.200	2227.867	1376.078	849.969	525.001	324.278	200.297	123.717	76.417	47.200	29.154	18.008	11.123	6.870	4.244	2.621	1.619	1.000
29.154	1376.078	849.969	525.001	324.278	200.297	123.717	76.417	47.200	29.154	18.008	11.123	6.870	4.244	2.621	1.619	1.000	0.618
18.008	849.969	525.001	324.278	200.297	123.717	76.417	47.200	29.154	18.008	11.123	6.870	4.244	2.621	1.619	1.000	0.618	0.382
11.123	525.001	324.278	200.297	123.717	76.417	47.200	29.154	18.008	11.123	6.870	4.244	2.621	1.619	1.000	0.618	0.382	0.236
6.870	324.278	200.297	123.717	76.417	47.200	29.154	18.008	11.123	6.870	4.244	2.621	1.619	1.000	0.618	0.382	0.236	0.146
4.244	200.297	123.717	76.417	47.200	29.154	18.008	11.123	6.870	4.244	2.621	1.619	1.000	0.618	0.382	0.236	0.146	0.090
2.621	123.717	76.417	47.200	29.154	18.008	11.123	6.870	4.244	2.621	1.619	1.000	0.618	0.382	0.236	0.146	0.090	0.056
1.619	76.417	47.200	29.154	18.008	11.123	6.870	4.244	2.621	1.619	1.000	0.618	0.382	0.236	0.146	0.090	0.056	0.034
1.000	47.200	29.154	18.008	11.123	6.870	4.244	2.621	1.619	1.000	0.618	0.382	0.236	0.146	0.090	0.056	0.034	0.021
0.618	29.154	18.008	11.123	6.870	4.244	2.621	1.619	1.000	0.618	0.382	0.236	0.146	0.090	0.056	0.034	0.021	0.013
0.382	18.008	11.123	6.870	4.244	2.621	1.619	1.000	0.618	0.382	0.236	0.146	0.090	0.056	0.034	0.021	0.013	0.008
0.236	11.123	6.870	4.244	2.621	1.619	1.000	0.618	0.382	0.236	0.146	0.090	0.056	0.034	0.021	0.013	0.008	0.005
0.146	6.870	4.244	2.621	1.619	1.000	0.618	0.382	0.236	0.146	0.090	0.056	0.034	0.021	0.013	0.008	0.005	0.003
0.090	4.244	2.621	1.619	1.000	0.618	0.382	0.236	0.146	0.090	0.056	0.034	0.021	0.013	0.008	0.005	0.003	0.002
0.056	2.621	1.619	1.000	0.618	0.382	0.236	0.146	0.090	0.056	0.034	0.021	0.013	0.008	0.005	0.003	0.002	0.001
0.034	1.619	1.000	0.618	0.382	0.236	0.146	0.090	0.056	0.034	0.021	0.013	0.008	0.005	0.003	0.002	0.001	0.001
0.021	1.000	0.618	0.382	0.236	0.146	0.090	0.056	0.034	0.021	0.013	0.008	0.005	0.003	0.002	0.001	0.001	0.000
Color Codes	Within the limits and defined by the scale							Within the limits of the scale					Outside the limits of the scale				

Appendix B. Computational Experiment Results

Table B 1 Performance measures for Balanced Scale

n	CR Limit	Measured Scale Lower CR	Fundamental Scale Lower CR	Equal CR	Fundamental Scale Inconsistent	Measured Scale Inconsistent	Both Scales Consistent	Same Best Chosen	Same Worst Chosen	Kendall's Tau Mean	Tau Standard Deviation
3x3	0.150	62.50%	37.40%	0.10%	29.80%	16.60%	53.60%	96.40%	96.30%	0.9508	0.1736
	0.100	62.40%	37.60%	0.00%	34.10%	21.20%	44.70%	96.70%	97.20%	0.9576	0.1653
	0.050	61.80%	38.20%	0.00%	54.80%	23.30%	21.90%	100.00%	99.70%	0.9974	0.0378
4x4	0.150	63.40%	36.60%	0.00%	30.20%	20.80%	49.00%	91.50%	90.60%	0.9107	0.1632
	0.100	70.30%	29.70%	0.00%	45.20%	18.40%	36.40%	90.90%	93.00%	0.9205	0.1547
	0.050	69.80%	30.20%	0.00%	58.50%	24.40%	17.10%	93.90%	95.20%	0.9442	0.1360
5x5	0.150	70.20%	29.80%	0.00%	46.70%	20.00%	33.30%	88.20%	88.00%	0.9028	0.1383
	0.100	70.20%	29.80%	0.00%	53.60%	21.90%	24.50%	90.80%	89.30%	0.9226	0.1232
	0.050	76.80%	23.20%	0.00%	69.90%	20.40%	9.70%	92.90%	92.60%	0.9437	0.0992
6x6	0.150	75.50%	24.50%	0.00%	62.80%	17.70%	19.50%	88.10%	89.70%	0.9056	0.1076
	0.100	82.60%	17.40%	0.00%	73.10%	13.00%	13.90%	87.90%	89.70%	0.9126	0.1055
	0.050	85.50%	14.50%	0.00%	84.00%	12.20%	3.80%	90.70%	92.30%	0.9392	0.0881
7x7	0.150	84.20%	15.80%	0.00%	75.30%	11.60%	13.10%	86.70%	87.00%	0.9029	0.0904
	0.100	90.70%	9.30%	0.00%	85.10%	7.30%	7.60%	86.50%	88.30%	0.9161	0.0896

Table B 2 Performance measures for Geometric Scale

n	CR Limit	Measured Scale Lower CR	Fundamental Scale Lower CR	Equal CR	Fundamental Scale Inconsistent	Measured Scale Inconsistent	Both Scales Consistent	Same Best Chosen	Same Worst Chosen	Kendall's Tau Mean	Tau Standard Deviation
3x3	0.150	76.70%	23.30%	0.00%	35.30%	12.70%	52.00%	97.90%	97.30%	0.9665	0.1432
	0.100	72.60%	27.40%	0.00%	36.00%	19.50%	44.50%	99.20%	98.50%	0.9821	0.1019
	0.050	69.80%	30.20%	0.00%	49.60%	23.70%	26.70%	100.00%	99.70%	0.9969	0.0391
4x4	0.150	97.00%	3.00%	0.00%	65.10%	0.10%	34.80%	90.20%	92.90%	0.9125	0.1645
	0.100	96.30%	3.70%	0.00%	69.30%	0.70%	30.00%	91.10%	93.10%	0.9207	0.1548
	0.050	94.50%	5.50%	0.00%	80.90%	1.40%	17.70%	94.60%	95.40%	0.9467	0.1324
5x5	0.150	99.90%	0.10%	0.00%	85.10%	0.00%	14.90%	87.50%	90.50%	0.8964	0.1400
	0.100	100.00%	0.00%	0.00%	91.60%	0.00%	8.40%	88.40%	89.80%	0.9048	0.1274
	0.050	99.10%	0.90%	0.00%	96.60%	0.10%	3.30%	89.80%	91.40%	0.9230	0.1133
6x6	0.150	100.00%	0.00%	0.00%	96.30%	0.00%	3.70%	84.30%	89.70%	0.8835	0.1229
	0.100	100.00%	0.00%	0.00%	97.90%	0.00%	2.10%	87.50%	87.90%	0.8989	0.1095
	0.050	100.00%	0.00%	0.00%	99.90%	0.00%	0.10%	88.10%	88.30%	0.9137	0.1013
7x7	0.150	100.00%	0.00%	0.00%	99.40%	0.00%	0.60%	82.60%	86.80%	0.8765	0.1112
	0.100	100.00%	0.00%	0.00%	99.80%	0.00%	0.20%	83.00%	88.30%	0.8959	0.0989

Table B 3 Performance measures for Inverse Linear Scale

n	CR Limit	Measured Scale Lower CR	Fundamental Scale Lower CR	Equal CR	Fundamental Scale Inconsistent	Measured Scale Inconsistent	Both Scales Consistent	Same Best Chosen	Same Worst Chosen	Kendall's Tau Mean	Tau Standard Deviation
3x3	0.150	62.30%	37.60%	0.10%	33.90%	21.40%	44.70%	95.60%	95.40%	0.9385	0.1911
	0.100	63.50%	36.50%	0.00%	41.60%	22.20%	36.20%	95.90%	96.30%	0.9457	0.1857
	0.050	63.10%	36.80%	0.10%	55.90%	24.90%	19.20%	97.90%	98.10%	0.9706	0.1405
4x4	0.150	67.10%	32.90%	0.00%	46.60%	24.50%	28.90%	86.20%	87.20%	0.8715	0.1911
	0.100	75.00%	25.00%	0.00%	57.00%	18.80%	24.20%	90.20%	89.50%	0.8955	0.1720
	0.050	71.30%	28.70%	0.00%	63.10%	23.60%	13.30%	93.30%	93.50%	0.9342	0.1475
5x5	0.150	76.40%	23.60%	0.00%	65.50%	18.50%	16.00%	83.90%	85.40%	0.8525	0.1728
	0.100	81.10%	18.90%	0.00%	74.10%	16.10%	9.80%	82.80%	88.70%	0.8822	0.1426
	0.050	85.10%	14.90%	0.00%	82.70%	13.60%	3.70%	86.30%	87.70%	0.9025	0.1277
6x6	0.150	87.00%	13.00%	0.00%	81.90%	11.00%	7.10%	83.20%	81.80%	0.8557	0.1335
	0.100	90.40%	9.60%	0.00%	88.00%	8.70%	3.30%	83.90%	85.30%	0.8740	0.1226
	0.050	94.10%	5.90%	0.00%	93.80%	5.40%	0.80%	86.60%	87.60%	0.9031	0.1134
7x7	0.150	95.70%	4.30%	0.00%	94.10%	3.70%	2.20%	80.10%	81.00%	0.8587	0.1096
	0.100	97.20%	2.80%	0.00%	96.80%	2.60%	0.60%	84.60%	83.80%	0.8726	0.1102

Table B 4 Performance measures for Logarithmic Scale

n	CR Limit	Measured Scale Lower CR	Fundamental Scale Lower CR	Equal CR	Fundamental Scale Inconsistent	Measured Scale Inconsistent	Both Scales Consistent	Same Best Chosen	Same Worst Chosen	Kendall's Tau Mean	Tau Standard Deviation
3x3	0.150	27.70%	72.20%	0.10%	4.30%	14.70%	81.00%	99.40%	100.00%	0.9951	0.0530
	0.100	34.50%	65.50%	0.00%	6.50%	20.90%	72.60%	99.60%	100.00%	0.9953	0.0461
	0.050	44.20%	55.80%	0.00%	13.70%	17.70%	68.60%	99.40%	99.70%	0.9912	0.0666
4x4	0.150	3.10%	96.90%	0.00%	0.00%	32.10%	67.90%	99.60%	98.90%	0.9912	0.0531
	0.100	5.10%	94.90%	0.00%	0.20%	42.90%	56.90%	99.30%	99.30%	0.9928	0.0471
	0.050	11.50%	88.50%	0.00%	0.50%	46.20%	53.30%	99.40%	99.90%	0.9948	0.0397
5x5	0.150	0.30%	99.70%	0.00%	0.00%	53.20%	46.80%	98.20%	97.60%	0.9832	0.0552
	0.100	0.60%	99.40%	0.00%	0.00%	60.40%	39.60%	98.10%	98.50%	0.9875	0.0491
	0.050	1.30%	98.70%	0.00%	0.10%	69.40%	30.50%	99.00%	99.10%	0.9927	0.0380
6x6	0.150	0.00%	100.00%	0.00%	0.00%	67.50%	32.50%	96.80%	96.90%	0.9797	0.0525
	0.100	0.00%	100.00%	0.00%	0.00%	78.00%	22.00%	97.70%	97.70%	0.9845	0.0467
	0.050	0.40%	99.60%	0.00%	0.10%	84.30%	15.60%	98.20%	97.60%	0.9886	0.0395
7x7	0.150	0.00%	100.00%	0.00%	0.00%	83.80%	16.20%	96.40%	97.50%	0.9806	0.0414
	0.100	0.00%	100.00%	0.00%	0.00%	90.80%	9.20%	98.80%	97.00%	0.9842	0.0372

Table B 5 Performance measures for Power Scale

n	CR Limit	Measured Scale Lower CR	Fundamental Scale Lower CR	Equal CR	Fundamental Scale Inconsistent	Measured Scale Inconsistent	Both Scales Consistent	Same Best Chosen	Same Worst Chosen	Kendall's Tau Mean	Tau Standard Deviation
3x3	0.150	98.60%	1.30%	0.10%	24.20%	0.80%	75.00%	99.40%	99.80%	0.9934	0.0612
	0.100	98.30%	1.70%	0.00%	27.30%	0.40%	72.30%	99.60%	99.90%	0.9963	0.0477
	0.050	96.80%	3.00%	0.20%	39.50%	0.90%	59.60%	99.30%	99.20%	0.9891	0.0820
4x4	0.150	100.00%	0.00%	0.00%	57.70%	0.00%	42.30%	99.50%	98.70%	0.9912	0.0532
	0.100	99.90%	0.00%	0.10%	58.80%	0.00%	41.20%	98.90%	98.90%	0.9867	0.0664
	0.050	100.00%	0.00%	0.00%	61.50%	0.00%	38.50%	99.50%	99.00%	0.9943	0.0432
5x5	0.150	100.00%	0.00%	0.00%	74.70%	0.00%	25.30%	98.30%	97.50%	0.9785	0.0668
	0.100	100.00%	0.00%	0.00%	83.30%	0.00%	16.70%	98.30%	97.80%	0.9814	0.0581
	0.050	100.00%	0.00%	0.00%	87.00%	0.00%	13.00%	98.30%	97.50%	0.9862	0.0523
6x6	0.150	100.00%	0.00%	0.00%	88.70%	0.00%	11.30%	97.30%	95.80%	0.9704	0.0649
	0.100	100.00%	0.00%	0.00%	94.90%	0.00%	5.10%	97.10%	96.60%	0.9748	0.0577
	0.050	100.00%	0.00%	0.00%	96.50%	0.00%	3.50%	97.30%	98.40%	0.9862	0.0406
7x7	0.150	100.00%	0.00%	0.00%	97.30%	0.00%	2.70%	94.30%	95.70%	0.9665	0.0562
	0.100	100.00%	0.00%	0.00%	98.30%	0.00%	1.70%	94.80%	96.70%	0.9738	0.0483

Table B 6 Performance measures for Root Square Scale

n	CR Limit	Measured Scale Lower CR	Fundamental Scale Lower CR	Equal CR	Fundamental Scale Inconsistent	Measured Scale Inconsistent	Both Scales Consistent	Same Best Chosen	Same Worst Chosen	Kendall's Tau Mean	Tau Standard Deviation
3x3	0.150	1.40%	98.30%	0.30%	3.10%	3.50%	93.40%	99.20%	99.70%	0.9908	0.0717
	0.100	1.00%	98.70%	0.30%	2.90%	8.60%	88.50%	99.50%	99.30%	0.9902	0.0747
	0.050	1.10%	98.60%	0.30%	8.90%	0.70%	90.40%	99.00%	99.40%	0.9844	0.0882
4x4	0.150	0.00%	100.00%	0.00%	0.00%	24.90%	75.10%	99.50%	99.50%	0.9954	0.0380
	0.100	0.00%	100.00%	0.00%	0.00%	28.00%	72.00%	99.60%	99.50%	0.9957	0.0367
	0.050	0.00%	100.00%	0.00%	0.00%	27.90%	72.10%	99.80%	99.70%	0.9961	0.0313
5x5	0.150	0.00%	100.00%	0.00%	0.00%	45.40%	54.60%	99.10%	99.10%	0.9901	0.0432
	0.100	0.00%	100.00%	0.00%	0.00%	51.50%	48.50%	99.00%	98.30%	0.9910	0.0424
	0.050	0.00%	100.00%	0.00%	0.00%	52.90%	47.10%	99.60%	99.10%	0.9955	0.0294
6x6	0.150	0.00%	100.00%	0.00%	0.00%	65.30%	34.70%	97.40%	98.50%	0.9848	0.0460
	0.100	0.00%	100.00%	0.00%	0.00%	67.50%	32.50%	98.40%	98.80%	0.9897	0.0370
	0.050	0.00%	100.00%	0.00%	0.00%	70.90%	29.10%	99.00%	99.20%	0.9949	0.0262
7x7	0.150	0.00%	100.00%	0.00%	0.00%	76.10%	23.90%	97.60%	95.80%	0.9855	0.0365
	0.100	0.00%	100.00%	0.00%	0.00%	82.60%	17.40%	97.00%	98.30%	0.9890	0.0321

Table B 7 Performance measures for FSBES

n	CR Limit	Measured Scale Lower CR	Fundamental Scale Lower CR	Equal CR	Fundamental Scale Inconsistent	Measured Scale Inconsistent	Both Scales Consistent	Same Best Chosen	Same Worst Chosen	Kendall's Tau Mean	Tau Standard Deviation
3x3	0.150	67.00%	32.90%	0.10%	25.00%	15.40%	59.60%	96.40%	97.30%	0.9555	0.1643
	0.100	68.90%	31.10%	0.00%	41.00%	11.90%	47.10%	97.40%	97.00%	0.9608	0.1540
	0.050	65.50%	34.40%	0.10%	54.70%	17.40%	27.90%	100.00%	99.80%	0.9961	0.0367
4x4	0.150	85.20%	14.80%	0.00%	39.90%	4.00%	56.10%	91.90%	92.30%	0.9182	0.1585
	0.100	83.10%	16.90%	0.00%	50.30%	5.60%	44.10%	93.10%	95.20%	0.9407	0.1341
	0.050	83.00%	17.00%	0.00%	62.90%	8.10%	29.00%	94.80%	95.80%	0.9517	0.1205
5x5	0.150	95.10%	4.90%	0.00%	66.90%	1.40%	31.70%	91.50%	91.60%	0.9207	0.1229
	0.100	95.60%	4.40%	0.00%	76.00%	1.20%	22.80%	90.50%	90.90%	0.9263	0.1164
	0.050	94.10%	5.90%	0.00%	83.50%	3.10%	13.40%	92.60%	94.40%	0.9431	0.1019
6x6	0.150	98.50%	1.50%	0.00%	83.20%	0.70%	16.10%	87.10%	91.30%	0.9164	0.1009
	0.100	98.70%	1.30%	0.00%	90.00%	0.30%	9.70%	88.40%	90.80%	0.9212	0.0991
	0.050	99.10%	0.90%	0.00%	96.50%	0.40%	3.10%	91.00%	90.50%	0.9337	0.0885
7x7	0.150	99.40%	0.60%	0.00%	94.00%	0.10%	5.90%	86.80%	90.50%	0.9136	0.0919
	0.100	99.90%	0.10%	0.00%	97.70%	0.00%	2.30%	90.40%	90.40%	0.9282	0.0826