B.C.YILDIRIM
METU 2019

A NEW PAIRWISE COMPARISON SCALE FOR ANALYTIC HIERARCHY PROCESS

BOĞAÇ CAN YILDIRIM

JANUARY 2019

A NEW PAIRWISE COMPARISON SCALE FOR ANALYTIC HIERARCHY PROCESS

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF SOCIAL SCIENCES OF MIDDLE EAST TECHNICAL UNIVERSITY

BY

BOĞAÇ CAN YILDIRIM

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF BUSINESS ADMINISTRATION
IN
THE DEPARTMENT OF BUSINESS ADMINISTRATION

JANUARY 2019

Approval of the Graduate School of Social Sciences

Prof. Dr. Tülin GENÇÖZ
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Business Administration.

Prof. Dr. Nuray GÜNER
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Business Administration.

Dr. M. Sinan GÖNÜL Co-Supervisor	Asst. Prof. Dr. Gülşah KARAKAYA Supervisor

Examining Committee Members

Prof. Dr. Uğur SOYTAŞ (METU, BA)

Asst. Prof. Dr. Gülşah KARAKAYA (METU, BA)

Asst. Prof. Dr. Özlem KARSU (Bilkent Uni., IE)

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Boğaç Can Yıldırım

Signature :

ABSTRACT

A NEW PAIRWISE COMPARISON SCALE FOR
ANALYTIC HIERARCHY PPROCESS

Yıldırım, Boğaç Can
MBA, Department of Business Administration
Supervisor: Asst. Prof. Dr. Gülşah Karakaya
Co-Supervisor: Dr. M Sinan Gönül

January 2019, 114 pages

One of the most significant difficulties in daily life or business decision problems is that they often involve multiple criteria, alternatives and/or stakeholders. Analytic Hierarchy Process (AHP) is one of the most widely used multi-criteria decision making tools in such problems. Despite its wide acceptance due to its systematic and simple 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 comparison of alternatives, which are then converted to numerical scores with a one-to-one mapping between the verbal comparisons and and a numerical scale. The choice of numerical scale affects one of the most important characteristics of pairwise comparisons, which is named as “consistency”. This study includes the comparison of the most widely used numerical pairwise comparison scale (Fundamental Scale) with other main numerical scales that have been suggested since the first foundation of AHP (Saaty, 1980). In the comparison procedure, the limitations of Fundamental Scale are identified, a new scale is proposed considering these limitations, and characteristics of all numerical pairwise comparison scales are analyzed. These analyses are tested with extensive simulations. All numerical scales are evaluated on an example decision making problem. Lastly, the advantages and disadvantages of the numerical scales are presented.

Keywords: Analytic Hierarchy Process (AHP), Pairwise comparison scale, Consistency, Simulation

ÖZ

ANALİTİK HİYERARŞİ SÜRECİ İÇİN
YENİ İKİLİ KARŞILAŞTIRMA ÖLÇEĞİ

Yıldırım, Boğaç Can
Yüksek Lisans, İşletme Bölümü
Tez Yöneticisi: Dr. Öğr. Üyesi Gülşah Karakaya
Ortak Tez Yöneticisi: Dr. M. Sinan Gönül

Ocak 2019, 114 sayfa

Günlük hayatta ya da çalışma ortamında karşılaşılan karar verme problemlerinde yaşanan en önemli zorluklardan biri, birden fazla ölçüt, seçenek ve/veya paydaşı içermeleridir. Bu tarz karar verme problemlerinde sıklıkla kullanılan çok kriterli karar verme yöntemlerinden biri olan Analitik Hiyerarşi Süreci (AHS), karar vericiler tarafından tercih edilmesini sağlayan sistematik ve anlaşılır yapısına karşın, yöntemin ana uygulamalarından biri olan ikili karşılaştırmalarda en sık kullanılan sayısal ölçek açılarından yetersizlikler de içermektedir. AHS, alternatiflerin sözel olarak karşılaştırılması ve bu sözel karşılaştırmaların bir sayısal ölçeğe göre sayısal puanlara çevrilmesi prensibiyle uygulanır. AHS'de kullanılan sayısal ölçek aynı zamanda yöntemin "tutarlılık" adı verilen en önemli karakterinden biri üzerinde etkilidir.Bu çalışma, AHS'de en sık kullanılan sayısal ikili karşılaştırma ölçeği olan Temel Ölçek ile AHS'nin öne sürülüşünden beri (Saaty, 1980) önerilmiş olan diğer ana sayısal ikili karşılaştırma ölçeklerinin karşılaştırmasını içermektedir. Bu karşılaştırmalar sırasında öncelikle Temel Ölçek'in yetersizlikleri belirtilmiş; bu yetersizlikleri gidereceği düşünülen yeni bir ölçek önerilmiş; sonrasında ise bütün ölçeklerin genel özellikleri analiz edilmiştir. Bu analizler yapılan geniş kapsamlı simülasyonlarla test edilmiş ve daha sonra bütün ölçekler bir örnek karar verme uygulaması üzerinde değerlendirilerek AHS için fayda ve sakıncaları yorumlanmıştır.

Anahtar Kelimeler: Analitik Hiyerarşi Süreci (AHS), İkili karşılaştırma ölçeği, Tutarlılık, Benzetim

This thesis is dedicated to my family, my love, and the people who supported me all the way through.

ACKNOWLEDGEMENTS

I would first like to express my sincere thanks to my dear advisor Asst. Prof. Dr. Gülşah KARAKAYA for her unlimited energy, patience, and her valuable guidance throughout this study.

I would also like to thank my co-advisor Assoc. Prof. Dr. M. Sinan GÖNÜL for his insights and guidance when I needed help.

I would like to specially thank to my family and my love for their patience and support during the entire process. It would be much more difficult if they hadn’t been there for me everytime needed.

TABLE OF CONTENTS

PLAGIARISM iii
ABSTRACT iv
ÖZ vi
DEDICATION viii
ACKNOWLEDGEMENTS ix
TABLE OF CONTENTS x
LIST OF TABLES xii
LIST OF FIGURES xv
LIST OF ABBREVIATIONS xvi
CHAPTER
1. INTRODUCTION 1
2. LITERATURE REVIEW 5
2.1. Overview of Analytic Hierarchy Process and its Application Areas 5
2.2. Axioms of Analytic Hierarchy Process 7
2.3. The Procedure of Analytic Hierarchy Process 9
2.3.1. Definition and Hierarchical Representation of a Decision Problem 10
2.3.2. Pairwise Comparisons 13
2.3.3. Weight Derivation 21
2.3.3.1. Eigenvalue Method (EVM) 28
2.3.3.2. Logarithmic Least Squares Method (LLSM) 29
2.3.4. Consistency Measurement 30
2.3.5. Aggregation of the Local Priorities 37
2.3.5.1. Additive Aggregation 37
2.3.5.2. Multiplicative Aggregation 38
3. LIMITATIONS OF EXISTING SCALES 40

4. A NEW PAIRWISE COMPARISON SCALE BASED ON FIBONACCI SEQUENCE 45
4.1. Scale-Based Inconsistency in the Existing Scales and Exponential Scale 49
5. SIMULATION RESULTS 52
5.1. Comparison of All Scales with Fundamental Scale 52
5.2. Performance Measures Used in Simulations 56
5.3. Simulation Steps and Results 57
6. COMPARISON OF PAIRWISE COMPARISON SCALES ON A SAMPLE DECISION PROBLEM 67
7. CONCLUSIONS AND FINAL REMARKS 74
REFERENCES 80
APPENDICES
A. SCALE BASED CONSISTENCY ASSESSMENTS 90
B. SIMULATION RESULTS 98
C. TURKISH SUMMARY/TÜRKÇE ÖZET 105
D. TEZ İZİN FORMU / THESIS PERMISSION FORM 114

LIST OF TABLES

Table 1 Pairwise comparison scale based on Saaty's Fundamental Scale 14
Table 2 Empty pairwise comparison matrix for the sample decision problem 15
Table 3 Main diagonal elements of the pairwise comparison matrix 16
Table 4 Reciprocal scores of compared elements 16
Table 5 Complete pairwise comparison matrix 17
Table 6 Pairwise comparison scales 18
Table 7 Level-2 Pairwise comparison matrix 22
Table 8 Level-2 Pairwise comparison matrix - Column summation 22
Table 9 Level-2 Pairwise comparison matrix - Column normalization 23
Table 10 Level-2 Pairwise comparison matrix - Row averaging 23
Table 11 Calculated weights for level-3 of the hierarchy 24
Table 12 Global criteria weights 25
Table 13 Local weights of alternatives in each lowest-level sub-criterion 26
Table 14 RI(n) values from various authors 33
Table 15 Global sub-criteria weights and Phone1's respective performances on each sub-criterion 37
Table 16 Aggregated scores of all alternatives by additive and multiplicative aggregation approaches 39
Table 17 Values of (a_{ij} * a_{jk}) in Fundamental Scale 43
Table 18 Adjusted one-to-one mapping between the verbal scale and numerical grades 46
Table 19 Approximate numerical pairwise comparison values (Initial) 47
Table 20 Approximate numerical pairwise comparison values (Final) 49
Table 21 Possible paired-combination multiplication results within limits of the scale 50

Table 22 Possible paired-combination multiplication results defined by the scale ... 51
Table 23 Numerical values for each scale, generated by the simulation ... 53
Table 24 Previously calculated RI values for each scale ... 54
Table 25 RI values calculated based on our simulation ... 55
Table 26 Performance Measurement of Different Scales ... 59
Table 27 Average performances of scales in terms of performance measures (all matrix sizes, CR limit=0.10) ... 63
Table 28 Inconsistent PCM (S) for the decision of buying a house for a family ... 67
Table 29 Consistent PCM (S') for the decision of buying a house for a family ... 68
Table 30 Weights and consistency measures for different pairwise comparison scales on inconsistent matrix S ... 69
Table 31 Weights and consistency measures for different pairwise comparison scales on consistent matrix S' ... 70
Table 32 Percent changes between S and S' cases ... 72
Table 33 Classification of judgement scales based on consistency and allocation of weights ... 73
Table A 1 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Balanced Scale ... 90
Table A 2 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Exponential Scale ... 91
Table A 3 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in Fundamental Scale ... 92
Table A 4 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Geometric Scale ... 93
Table A 5 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Inverse Linear Scale ... 94
Table A 6 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Logarithmic Scale ... 95
Table A 7 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Power Scale ... 96

Table A 8 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Root Square Scale 97
Table B 1 Performance measures for Balanced Scale 98
Table B 2 Performance measures for Exponential Scale 99
Table B 3 Performance measures for Geometric Scale 100
Table B 4 Performance measures for Inverse Linear Scale 101
Table B 5 Performance measures for Logarithmic Scale 102
Table B 6 Performance measures for Power Scale 103
Table B 7 Performance measures for Root Square Scale 104

LIST OF FIGURES

Figure 1 The example used by Saaty and Vargas (2012) 8
Figure 2 Exemplary hierarchy for an automobile selection problem 8
Figure 3 Overall goal of the decision problem 11
Figure 4 Main criteria of the decision problem 11
Figure 5 Sub-criteria of the decision problem 12
Figure 6 Complete decision hierarchy the "Mobile Phone Selection" problem 13
Figure 7 Local weights of each criterion in the hierarchy 25
Figure 8 General classification of weight derivation methods 27
Figure 9 Classification of the frequently used weight derivation methods 27
Figure 10 Exponential fit to the Fibonacci numbers based on the adjusted numerical grades 46
Figure 11 Equation and goodness of exponential fit 47
Figure 12 Comparison of scales based on the CR of generated matrices 61
Figure 13 Comparison of scales based on same best selection 61
Figure 14 Comparison of scales based on same worst selection 62
Figure 15 Comparison of scales based on Kendall's Tau performance 62
Figure 16 Comparison of scales based on estimated performance scores (CR=0.10) 65
Figure 17 Comparison of scales based on estimated performance scores (CR=0.15) 65
Figure 18 Comparison of scales based on estimated performance scores (CR=0.05) 66

LIST OF ABBREVIATIONS

AHP	Analytic Hierarchy Process
DM	Decision Maker
EVM	Eigenvalue Method
GCI	Geometric Consistency Index
LLSM	Logarithmic Least Squares Method
PCM	Pairwise Comparison Matrix
RGMM	Row Geometric Mean Method

CHAPTER 1

INTRODUCTION

A world with decisions only based on a single criterion would be an easy environment to overcome decision problems. It would simply be the choice of the alternative, which provides the best performance on the single decision criterion. For instance, if a decision maker (DM) were to buy a car considering only its engine capacity, indeed, he/she would be overlooking many other features of the car, such as its passenger capacity, comfort, and safety. Conversely, in real life, the decision problems which require a single decision criterion are very rare compared to those which require multiple decision criteria. DMs often confront with situations where they need to make trade-offs, i.e., to stretch some criteria or even waive some of them in order to have better performance on others. In this regard, humans need to evaluate each decision criterion considering the other decision criteria in order to decide which alternative fits best to their expectations, even if an alternative performs the worst with respect to one (or multiple) criterion.

Management decisions are often complex as they involve multiple objectives (such as profit maximization, cost minimization) and many decision criteria. In many cases, there are more than one stakeholder, which makes the decision problems even more complex as the DM may have to ensure some certain requirements are fulfilled (i.e. safety requirements by laws, public satisfaction etc.). While “fast and frugal heuristics” (Gigerenzer, Gerd; Todd, P. M.; The ABC Research Group, 1999) work very well in some situations, solid analyses based on systematic approaches are more preferable in managerial decision making.

Having grown as a part of the Operational Research, Multi-Criteria Decision Making (MCDM) is concerned with providing verbal and/or computational tools to support a DM while he/she evaluates two or more alternatives with respect to two or more related/unrelated decision criteria. MCDM idea aims to propose an "optimal" or "good enough" solution to decision problems, as it is nearly impossible to have the best results in every criteria of the DM.

Numerous MCDM techniques have been proposed by many researchers as systematic approaches to multi-criteria decision problems. Some of the most well-known teqhniques can be listed as:

- Analytic Hierarchy Process (AHP) (Saaty, 1980)
- The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) (Hwang and Yoon, 1981)
- Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) (Brans, 1982)
- Simple Multi Attribute Rating Technique (SMART) (von Winterfeldt and Edwards, 1986)
- ELimination Et Choix Traduisant la Realité (ELECTRE) (Roy and Bouyssou, 1993)
- Measuring Attractiveness Through a Categorical-Based Evaluation Technique (MACBETH) (Bana e Costa et al., 2012)

Although there are many other techniques, none of them can be regarded as the "best" MCDM technique that is superior to the others in every aspect. Indeed, several studies (Badri, 2001; Macharis et al., 2004; Pirdashti et al., 2009; Amiri et al., 2009) employ hybrid use of techniques to provide improved suggestions to DMs.

Among these techniques, AHP attracts attention as it is the first one to systematically use the concept of "pairwise comparisons", which was suggested by Fechner (1860), and developed by Thurstone (1927). According to the concept of pairwise comparisons, elements constituting the complex decision problem should be compared in pairs in order not to exceed the cognitive capacity of human mind and obtain invalid results. Another strength of AHP is that it enables the DM to model complex decision problems by dividing them in simpler portions, and then, summarizing the results of these portions to come up with the best alternative based on his/her preferences. Moreover, AHP does not necessarily require a sophisticated software.

Its strengths led to its widespread use in many real life decision making applications. Being a widely used method also attracted the attention of the academy and made AHP one of the most studied MCDM techniques. These studies criticize AHP in various aspects. Some of them focus on the scale used in pairwise comparisons (Dong et al., 2008; Franek and Kresta, 2014), while some other studies elaborate on the weight extraction (Saaty and Hu, 1998; Dijkstra, 2013) and inconsistency measurement (Dodd et al., 1992; Davoodi, 2009). While some of these studies suggest novel ideas to improve AHP (Lin et al., 2013), some other studies simply compared and contrasted what have already been proposed or is being widely used (Ishizaka and Lusti, 2006; Franek and Kresta, 2014).

This study also reviews the literature, criticizes some parts of AHP, proposes a novel pairwise comparison scale with better consistency characteristics and introduces new performance measures, which have not yet been used in any previous research in AHP literature.

The rest of this study is organized as follows: Chapter 2 provides an overview of the Analytic Hierarchy Process and its application areas, defines its application procedure, and explains the fundamental axioms that constitute the basis of AHP methodology. In the rest of the Chapter 2, the procedure of AHP is elaborated on by detailed explanation of its main steps, namely “Definition and Hierarchical Representation of Decision Problem”, “Pairwise Comparisons”, “Weight Derivation”, “Consistency Measurement”, and “Aggregation of the Local Priorities”. Chapter 3 discusses the “Limitations of Existing Scales”. In Chapter 4, a new pairwise comparison scale is proposed in order to overcome the limitations analyzed in Chapter 3. Then, Chapter 5 presents the numerical analyses and simulation results comparing the proposed pairwise comparison scale and the existing ones with respect to related performance measures. Chapter 6 compares the scales through a sample decision problem. Finally, in Chapter 7, the findings of the analyses are discussed details and possible future study areas with conclusive remarks are presented.

CHAPTER 2

LITERATURE REVIEW

In this chapter, an overview of the Analytic Hierarchy Process (AHP) will be provided by introducing its application areas, defining its procedure and explaining the main axioms underlying the whole methodology.

2.1. Overview of Analytic Hierarchy Process and its Application Areas

Analytic Hierarchy Process (AHP) is a multi-criteria decision making (MCDM) method, originally developed by Thomas L. Saaty (1980). It was developed as a reaction to the finding that there is a miserable lack of common, easily understood and easy-to-implement methodology to enable the taking of the complex decisions. Since then, the simplicity and power of AHP has led to its widespread use across multiple domains in every part of the world (El Hefnawy and Mohammed, 2014). AHP has been used in business, government, social studies, R&D, defense and other domains involving decisions in which choice, prioritization or forecasting are needed (Bhushan and Rai, 2004). As indicated in the original study (Saaty, 1980), AHP is a theory of measurement through pairwise comparisons and relies on the judgements of experts to derive priority scales.

The importance of AHP comes from its wide applicability to real life problems without many assumptions and adjustments. Although there are various softwares to apply AHP, the entire process basically requires only one critical asset: a DM who is knowledgeable enough in the decision topic so that he/she can accurately compare the items involved in the decision making problem. Based on several studies (Vaidya and Kumar, 2006; El Hefnawy and Mohammed, 2014; Russo and Camanho, 2015; Schmidt et al., 2015) application types of AHP in real-life decision problems include, but are not limited to:

- Selection of an alternative from a given set of alternatives
- Evaluation and performance measurement of multiple alternatives
- Benefit-cost analysis
- Resource allocation
- Planning and development
- Priority determination and ranking of alternatives from the most to the least desirable
- Forecasting and outcome prediction

Based on the same studies and literature review, some application areas of AHP in real-life decision problems include, but are not limited to:

- Project management (Al-Harbi, 2001)
- Defense (Cheng, 1997)
- Mining (Ehie and Benjamin, 1993)
- Aerospace (Tavana, 2003)
- Healthcare (Schmidt et al., 2015)
- Environment management (Stirn and Grošelj, 2010)
- Property (Safian and Nawawi , 2011)
- Forecasting (Blaira et al., 2002)
- Engineering (Chen and Lin, 2003)
- Location selection (Takamura and Tone, 2003)
- Manufacturing (Abdi and Labib, 2003)

Another advantage of AHP is that it enables post-process evaluation of the DM's judgements through numerical analysis. This way, the method checks whether or not the DM is consistent in his/her pairwise comparisons.

2.2. Axioms of Analytic Hierarchy Process

Harker and Vargas (1987) note that Saaty (1986) has defined four axioms that constitute the base of AHP. These axioms are necessary to have a complete understanding of AHP, with the reasons of its simplicity and related drawbacks:

- Axiom 1 – Reciprocal Condition: Although pairwise comparison questions are asked verbally, AHP is a numerical tool. According to the first axiom, for any pair of compared elements, the intensity of preference of Element1 over Element2 is inversely related to the intensity of preference of Element2 over Element1. To simply restate, if Element1 is 5 times more preferable than Element2, then Element2 is 1/5 as desirable as Element1.

- Axiom 2 – Homogeneity: Saaty (1986) states that individuals are only capable of expressing meaningful intensities of preference if the elements are comparable. According to Saaty (1987), homogeneity is essential as the mind cannot compare widely disparate elements. Sagir Ozdemir (2005) notes that pairwise comparisons can be applied successfully to stimuli that are not too disparate in their magnitudes with respect to the possession of a certain attribute. The difficulty of comparison increases as the disparity between the compared elements increase. Saaty and Vargas (2012) exemplify this with the comparison of an unripe cherry tomato with an oblong water melon in terms of their volumes. Instead, a clustering of comparable-sized (homogenous) objects is proposed in order to create conceivable steps of comparisons. Then, these comparisons are used to obtain an approximation of the comparison by multiplication. The example is illustrated by Saaty and Vargas (2012) as shown in Figure 1.

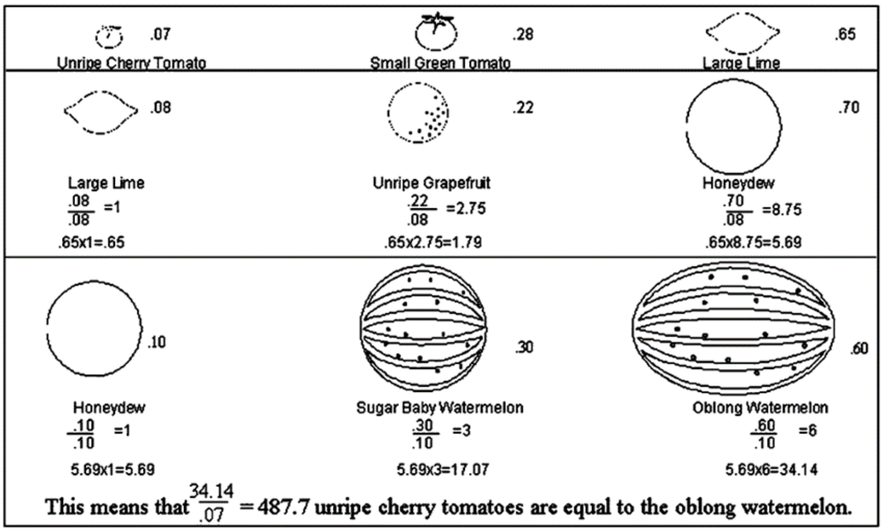

Figure 1 The example used by Saaty and Vargas (2012)

- <u>Axiom 3 – Dependence:</u> All sets of elements in the hierarchy should be compared in terms of the element on their immediately upper hierarchy. Consider the hierarchy given in Figure 2. For instance, if three different automobiles are to be compared in terms of trunk size (a sub-criteria under the "Physical Properties" main criteria), all of them should be compared with each other in terms of the trunk size, not in terms of another sub-criteria under "Physical Properties" or the "Physical Properties" itself, which is two levels above the compared elements (cars) in the hierarchy.

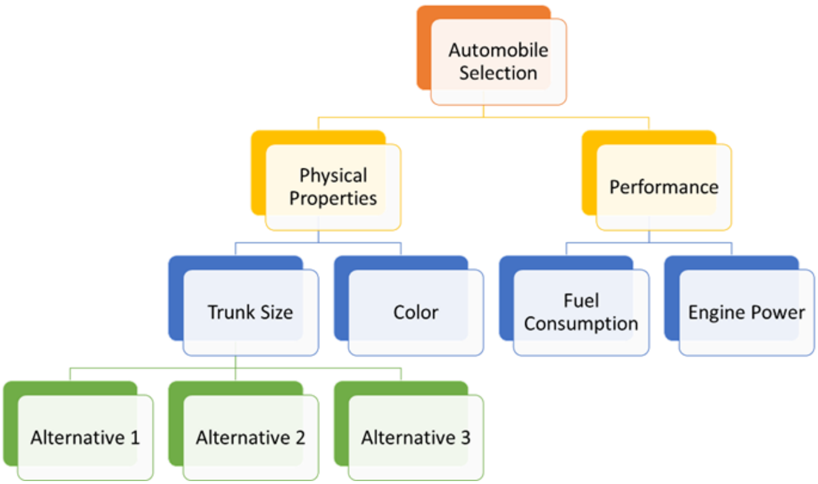

Figure 2 Exemplary hierarchy for an automobile selection problem

- Axiom 4 – Expectations: If an exact replica of an alternative is added to the comparison, the DM should either adjust the criteria by adding another criterion such as “the number of elements of certain type” or simply do not add that replica as the preference of A over B is the same for the preference of A’s replica (A’) over B.

2.3. The Procedure of Analytic Hierarchy Process

AHP procedure consists of five main steps:

- Definition and hierarchical representation of the decision problem: This step is basically the determination of the main aim of AHP application and division of a complex problem into smaller manageable parts, in which, elements involved in the problem can be compared in pairs.
- Pairwise comparisons: Each element involved in the problem are verbally compared with the other elements in its respective hierarchy level, and these comparisons are transcribed to numerical values using a one-to-one mapping between the verbal comparison scale and corresponding numerical scores. These numerical scores represent the “intensity of preference” of each element over others.
- Weight derivation: Based on the numerical scores obtained in the previous steps, a special matrix called “Pairwise Comparison Matrix” (PCM) is formed for all compared elements. Then, these numerical PCMs are evaluated using specific methods to extract the respective weights of each element within each PCM.
- Consistency measurement: The DM has already responded the verbal pairwise comparison questions. However, their internal validity (consistency) is not yet checked. Therefore, in this step, the DM’s consistency in answering the verbal pairwise comparison questions are checked using the numerical PCMs. If any

inconsistency is detected, then, the DM is asked to revise his/her pairwise comparisons as he/she might have made a mistake in judgements.

- Aggregation (synthesis) of the local priorities (weights): Weights calculated using PCMs are only the priorities (local priorities) based on those specific PCMs. Their contributions to the main aim of AHP application (overall goal), however, still need to be determined. Here, in this step, previously calculated local priorities are synthesized to the top level of the hierarchy.

The abovementioned steps will be discussed in details in the upcoming chapters of this study.

2.3.1. Definition and Hierarchical Representation of a Decision Problem

In multi-criteria decision problems, it is very unlikely, and mostly impossible, to select an alternative by evaluating all decision criteria and alternatives at once. In line with Simon (1955), Sagir Ozdemir (2005) notes that there is a limit in our ability to process information in making comparisons on a large set of elements. Therefore, a systematic division of the complex decision problem is necessary for a human to properly understand and process it. In parallel to this point of wiev, the first step in AHP is to divide a complex problem into manageable portions in order to properly deal with complexity. This division is made through the hierarchical representation of the complex decision problem.

According to El Hefnawy and Mohammed (2014), hierarchical representation of a decision problem has two basic advantages:

- It provides an overall view of the complex system of the situation, and
- It helps the DM assess the homogeneity of the issues in each level, so he/she can compare the items accurately.

Saaty (1994) suggests four main steps do build a proper hierarchical structure for AHP applications:

- Identify the overall goal
- Identify main criteria that must be satisfied to fulfill the overall goal
- If necessary, identify sub-criteria under each criterion
- Add alternatives under the lowest level criteria (the bottom of the hierarchical structure)

To illustrate the main steps of AHP, the hierarchy in a sample decision problem given in Işıklar and Büyüközkan (2007) will be used. Suppose a person wants to purchase a new mobile phone and she needs to make her decision between three different alternatives: Phone1, Phone2, and Phone3.

Based on Saaty's suggestion, the first step is to determine the overall goal of the decision problem. Generally, the problem definition involves its own overall goal. In our example, the overall goal is "Mobile Phone Selection", as shown in Figure 3.

Figure 3 Overall goal of the decision problem

The second step is to define the main criteria. According to Işıklar and Büyüközkan (2007), "Mobile Phone Selection" problem can be divided to two main criteria, namely "Product Related" and "User Related", as shown in Figure 4.

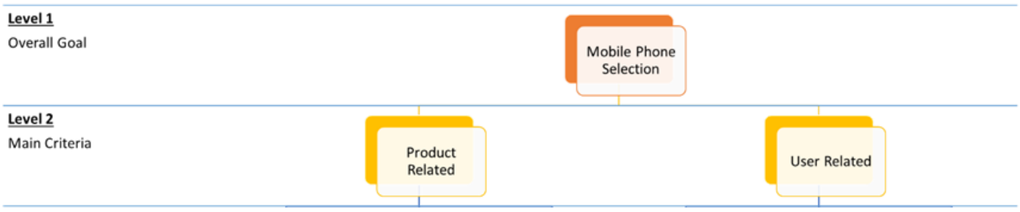

Figure 4 Main criteria of the decision problem

At this step, it is necessary to ask whether the main criteria are clear and detailed enough to make pairwise comparisons. Obviously, there may be many different product and user related decision criteria, and main criteria definitions are too general to be properly compared. Thus, it appears that the problem requires further division. The next step, then, is to determine the sub-criteria related to the product and the user, respectively. The study divides each main criterion to three sub-criteria, as shown in Figure 5.

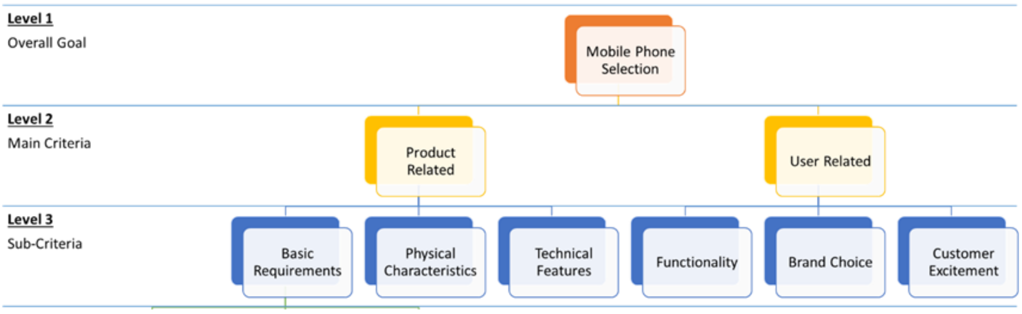

Figure 5 Sub-criteria of the decision problem

While dividing the problem into smaller parts, it is important not to divide the problem too much. When people compare items, they focus on the common attributes of those items and judge based on their differences in common attributes. If the problem is divided too much that the items do not have common attributes anymore, it becomes more difficult to compare the items in the same level of hierarchy. Considering our example, a three-level hierarchy for the overall goal and decision criteria are enough to represent the problem as detailed and comparable as possible. Thus, the next level in the hierarchy will be the alternatives, which will be compared during the application phases of our example. Complete hierarchy is shown in Figure 6. For the sake of illustrative simplicity, alternatives are shown only under “Basic Requirements” sub-criteria. However, it should be noted that all three alternatives are under each lowest-level criterion in AHP decision problems, as they share the lowest-level criteria as a common attribute.

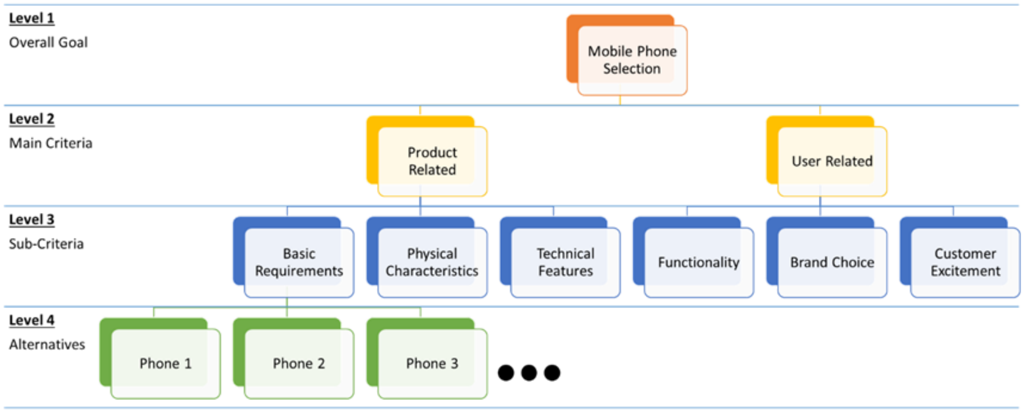

Figure 6 Complete decision hierarchy the "Mobile Phone Selection" problem

2.3.2. Pairwise Comparisons

The second step in AHP is the comparison of the criteria and alternatives within their hierarchy group at the same level of the hierarchy. Several studies (Alonso and Lamata, 2006; Dong et al., 2008) indicate that the concept of pairwise comparison originates from psychological study conducted by Fechner (1860), and developed by Thurstone (1927). According to a recent study (Fashoto et al., 2016), the pairwise comparison technique is commonly used to handle subjective and objective judgements in multi-criteria decision making. Sagir Ozdemir (2005) argues that humans are not sufficiently sensitive to make accurate changes in judgements on several elements simultaneously. Another study (El Hefnawy and Mohammed, 2014) suggests that the pairwise comparison process is strongly recommended by psychologists, as it is easier and more accurate to express opinion on only two alternatives than to do it simultaneously on all the alternatives.

Since the first invention of AHP by Saaty (1980), pairwise comparison scale has been one of the most widely discussed topics. AHP method suggests that DM verbally compares two elements at once, which share a common parent in the hierarchy. These verbal comparisons are converted to numerical values based on a one-to-one mapping between the set of discrete linguistic choices available to the DM and a discrete set of numbers, which represent the importance, or weight, of the previous linguistic choices (Triantaphyllou and Mann, 1995). The most commonly used one-to-one mapping is

Fundamental Scale (or as known as 1-9 Linear Scale) of Saaty (Franek and Kresta, 2014), which is given in Table 1.

Table 1 Pairwise comparison scale based on Saaty's Fundamental Scale

Intensity of Importance	**Definition**	**Explanation**
1	Equal importance	Two activities contribute equally to the objective
2	Weak or slight	
3	Moderate importance	Experience and judgement slightly favor one activity over another
4	Moderate plus	
5	Strong importance	Experience and judgement strongly favor one activity over another
6	Strong plus	
7	Very strong or demonstrated importance	An activity is favored very strongly over another; its dominance demonstrated in practice
8	Very, very strong	
9	Extreme importance	The evidence favoring one activity over another is of the highest possible order of affirmation
Reciprocals of above	If activity *i* has one of the above non-zero numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*	A reasonable assumption

Sources: Triantaphyllou and Mann (1995), Goepel (2018)

The mapping between the verbal and numerical scales is used for the purpose of recording each pairwise comparison between the elements. Once two elements are compared using the verbal scale, the verbal comparison is translated using the abovementioned one-to-one mapping to obtain a numerical score. The score for each comparison is noted on a special matrix called "Pairwise Comparison Matrix" or "Positive Reciprocal Matrix" as in Saaty (1980).

Recall the example given in "Definition and Hierarchical Representation of Decision Problem" section of this study. The DM's evaluation criteria at level 2 are divided into two main groups: "Product Related" and "User Related". These main groups are further divided into sub-criteria at level 3 of the decision hierarchy. At the last level, level 4, alternatives are listed for each sub-criterion. In our example in this section, only "Product Related" criterion is used. Based on these criteria, unevaluated pairwise comparison matrix is formed as in Table 2.

Table 2 Empty pairwise comparison matrix for the sample decision problem

Product Related Criteria	**Basic Requirements**	**Physical Characteristics**	**Technical Features**
Basic Requirements			
Physical Characteristics			
Technical Features			

The DM compares the row elements (i) with column elements (j), assesses verbally the importance of the row element (i) with respect to the column element (j). Then, these verbal comparisons are translated to numerical scores, using a pairwise comparison scale. Saaty's Fundamental Scale is used in this illustrative example.

Once comparison step is carried out, pairwise comparison matrix is filled with the respective numerical scores (a_{ij}). As all elements are equally important compared to themselves, it becomes $a_{ij} = 1$ for all "i=j", which means all numerical elements on the main diagonal of the pairwise comparison matrix is "1" (see Table 3). The DM does not need to make comparisons for these elements.

Table 3 Main diagonal elements of the pairwise comparison matrix

Product Related Criteria	**Basic Requirements**	**Physical Characteristics**	**Technical Features**
Basic Requirements	1		
Physical Characteristics		1	
Technical Features			1

Remaining elements, where i≠j, are compared by the DM. Suppose the DM compares cost and quality, resulting in a verbal statement of "Cost has very strong importance compared to comfort.". From the mapping in Table 1, it is retrieved that the corresponding score is "7" for a_{12}. The reciprocity axiom of AHP dictates "$a_{ij} = {}^{1}/_{a_{ji}}$" for all i, j. Therefore, it is obvious that "$a_{21} = 1/7$". Necessary entries yield the comparison matrix in Table 4.

Table 4 Reciprocal scores of compared elements

Product Related Criteria	**Basic Requirements**	**Physical Characteristics**	**Technical Features**
Basic Requirements	1	7	
Physical Characteristics	$\frac{1}{7}$	1	
Technical Features			1

Similarly, suppose the DM evaluated that "basic requirements are more important than technical features" and "technical features are more important than physical characteristics". Converting the verbal judgements to numerical scores yields the complete pairwise comparison matrix as in Table 5. It should be noted that it is enough for the DM to make the comparisons required to fill the upper triangle of the pairwise comparison matrix, as the other scores can be determined using the reciprocity axiom. Another point to note is that the upper triangle does not necessarily have to involve elements greater than or equal to "1". Recall that the DM judged that "safety is more important than comfort". Again, the reciprocity axiom dictates that if safety is more important than comfort, that is "$a_{32} = \frac{Technical\ Features}{Physical\ Characteristics} = 3$", then "$a_{23} = \frac{Physical\ Characteristics}{Technical\ Features} = \frac{1}{3}$".

Table 5 Complete pairwise comparison matrix

Product Related Criteria	**Basic Requirements**	**Physical Characteristics**	**Technical Features**
Basic Requirements	1	7	3
Physical Characteristics	$\frac{1}{7}$	1	$\frac{1}{3}$
Technical Features	$\frac{1}{3}$	3	1

Fundamental Scale was founded based on the psychophysical law of Weber-Fechner, and Saaty indicates that it was derived mathematically from stimulus response theory (Saaty, 1996). Saaty (1980; 1996) tested the 1-9 scale, and about twenty other scales to choose a suitable ratio scale for the pairwise comparisons in AHP. Based on their testing results, the 1-9 scale was accepted by AHP. Since then the 1-9 scale has become the most widely used ratio scale in AHP (Zhang et al., 2009).

However, the ratio scale used for pairwise comparisons is one of the most controversial areas in AHP literature. The values defined on a numerical scale are used to represent the relative importance between two compared objects in terms of a ratio (Ji and Jiang, 2003). Considering the fact that the scale used in AHP has a significant effect on the outcome of the process, the issue on numerical comparison scales has drawn the attention of the researchers. Despite being the most widely used ratio scale for quite a long time, Saaty’s Fundamental Scale has been criticized by researchers in many studies (see Table 6). Dodd et al. (1992) point out that the simplicity of the scoring method and the coarseness of the scale are inseparable in Saaty's method. Although the verbal comparison scale has not been a concern of the literature, several numerical scales have been proposed as an alternative for Saaty’s Fundamental Scale (Table 6).

Table 6 Pairwise comparison scales

Scale	Mathematical Description	Parameters (x)	Approximate Scale Values
Linear (Saaty, 1977)	x	$\{1,2,\ldots,9\}$	$1; 2; 3; 4; 5; 6; 7; 8; 9$
Power (Harker and Vargas, 1987)	x^2	$\{1,2,\ldots,9\}$	$1; 4; 9; 16; 25; 36; 49; 64; 81$
Root Square (Harker and Vargas, 1987)	$\sqrt{x}$	$\{1,2,\ldots,9\}$	$1; \sqrt{2}; \sqrt{3}; 2; \sqrt{5}; \sqrt{6}; \sqrt{7}; \sqrt{8}; 3$
Geometric (Lootsma, 1989)	2^{x-1}	$\{1,2,\ldots,9\}$	$1; 2; 4; 8; 16; 32; 64; 128; 256$
Inverse Linear (Ma and Zheng, 1991)	$\frac{9}{(10-x)}$	$\{1,2,\ldots,9\}$	$1; 1.13; 1.29; 1.5; 1.8; 2.25; 3; 4.5; 9$
Asymptotical (Dodd and Donegan, 1994)	$tanh^{-1}\left(\frac{\sqrt{3}(x-1)}{14}\right)$	$\{1,2,\ldots,9\}$	$0; 0.12; 0.24; 0.36; 0.46; 0.55; 0.63; 0.7; 0.76$
Balanced (Salo and Hämäläinen, 1997)	$\frac{x}{(1-x)}$	$\{0.5, 0.55, \ldots,0.9\}$	$1; 1.22; 1.5; 1.86; 2.33; 4; 5.67; 9$
Logarithmic (Ishizaka et al., 2010)	$log_2(x+1)$	$\{1,2,\ldots,9\}$	$1; 1.58; 2; 2.2; 2.58; 2.81; 3; 3.17; 3.32$

Source: Franek and Kresta (2014)

According to Franek and Kresta (Franek and Kresta, 2014), AHP needs ratio scales due to its pairwise comparison characteristics. It is claimed in several studies (Stevens, 1957; Stevens and Galanter, 1964) that ratio scales are appropriate means to elicit response stimuli. Harker and Vargas (1987) note that any ratio scale could be used in

AHP and the choice of a scale such as Saaty's Fundamental Scale is a result of experimental evidence.

The main motivation behind Saaty's Fundamental Scale is the aim to ensure that the scale does not exceed the capacity of the short- term memory and uses simple integer values (Ji and Jiang, 2003). Mazurek and Perzina (2017) emphasize that according to Miller (1956), a human brain is capable of processing only up to 7 pieces (chunks) of information at the same time. Therefore, they conclude that inconsistency increases with the increasing number of judgements. According to Saaty and Sagir Ozdemir (2003) the number of compared elements should not be more than 7, for the consideration of consistency of the information derived from the pairwise comparison matrix. Saaty (2001), on the other hand, states that the upper limit of the scale should not be greater than 9, for homogeneity axiom and consistency. Therefore, Saaty preferred a 9-point scale that has an upper limit of 9. For the sake of simplicity of understanding and application, in fact, there is no reason not to accept a linear 9-point numerical scale from 1 to 9.

The study conducted by Harker and Vargas (1987) has investigated a quadratic and a root square scale using a simple AHP example, and supported Saaty's Fundamental Scale. Yet, it has been criticized by Franek and Kresta (2014) that only one example is not enough to claim the superiority of Fundamental Scale. Another important critic on Fundamental Scale is that the scale has been supported by Saaty's empirical evidence, but it is not a transitive scale (Dong et al., 2008).

Lootsma (1989) argued that it is more convenient to use a geometric scale instead of Fundamental Scale. Lootsma's Geometric Scale rooted from psychological observations about stimulus perception and his conclusion from related studies (Lootsma, 1993; 1999) that "human beings follow exponential scales when they categorize an interval such as ranges of time, sound, and light intensities". That is to say, Lootsma's Geometric scale is founded based on assumptions or external observations.

Ma and Zheng (1991) criticized the linear characteristic of Fundamental Scale for numerical values between 1 and 9. They suggested a scale called “Inverse Linear Scale”, where the reciprocal elements (elements less than 1) of the scale are linear instead of the elements greater than “1” in Fundamental Scale. The basic consideration is to make the numerical values matched with the corresponding verbal expressions. This leads to the fact that the relation between $1/a_{ij}$ is linear rather than a_{ij} in Fundamental Scale.

Dodd et al. (1992) criticized Saaty’s Fundamental Scale remarking that the scale suffers from vagueness of definition and it is not closed under multiplication (boundary problem). Dodd and Donegan (1994) have proposed an asymptotic scale with the main motivation to avoid the boundary problem of Fundamental Scale. Asymptotic scale can approximately linearize Fundamental Scale in the neighborhood of 1, while further exponential mapping ensures that positive reciprocity holds for each value on the scale. 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 by the simulations.

According to Salo and Hämäläinen (1997), discretized ratio scales such as the 1-9 scale of AHP can be very helpful in preference elicitation; yet, they are problematic as they severely restrict the range and distribution of possible priority vectors. They point out that the integers on Fundamental Scale yield local weights, which are unequally dispersed. Therefore, they claim that elements preferentially close to each other lack sensitivity in pairwise comparison. Based on this approach, they proposed a “Balanced Scale” where the local weights are “evenly dispersed” over the weight range of 0.1-0.9, with an increment of 0.05 for each incremental step.

Ishizaka, Balkenborg and Kaplan (2010) emphasized that decision making scales in AHP should not tend to recommend extremes which are good in only one dimension. According to them, decision making is almost always making compromises.

Therefore, they suggested another pairwise comparison scale, Logarithmic Scale, which they claim offers more possibilities for the compromise alternative to be selected.

For all abovementioned scales, the linguistic expression of pairwise comparison judgements is kept as in Saaty's original setup (Saaty, 1977).

2.3.3. Weight Derivation

Once all required pairwise comparisons are made and pairwise comparison matrices with numerical ratio scores are generated, the respective importance of all compared elements need to be extracted. Just as pairwise comparison scales, weight (priority, importance) derivation methods have been one of the main discussion topics in AHP literature (Triantaphyllou and Mann, 1995).

Although numerous different methods have been suggested by researchers, according to Mazurek and Perzina (2017), weights of all elements (criteria and alternatives) are usually determined by Saaty's eigenvalue method (EVM) as the principal right eigenvector of the respective pairwise comparison matrix. Similarly, another study (El Hefnawy and Mohammed, 2014) strongly notes that the EVM is the most used method to derive weights in the vast majority of the applications of AHP. Another method, "Mean of Normalized Values" (or Rule of Thumb), is frequently used to approximate the EVM in many practical applications, as it is more practical than EVM and approximates well the overall weights in low inconsistency pairwise comparison matrices (Mu and Pereyra-Rojas, 2017). Therefore, for the sake of simplicity, Mean of Normalized Values (MNV) is used to illustrate how weights are derived in AHP in our example. The concept of "Consistency" will be discussed in details in "Consistency Measurement" section of this study.

Recall the "Mobile Phone Selection" example. The DM has a 4-level hierarchy. In AHP, each element is compared with the other elements that share the same immediate upper hierarchy element. That is, for our example, "Basic Requirements", "Physical Characteristics", and "Technical Features" are compared with each other as they all are immediately under "Product Related" criteria. Similarly, for the "User Related" criteria, "Functionality", "Brand Choice", and "Customer Excitement" are compared with each other. Once all elements at the same hierarchy level are grouped based on their upper hierarchy and compared as exemplified in "Pairwise Comparisons" section, elements on another level of hierarchy are compared with the same process.

Although there is no strict rule to compare hierarchy levels from top to bottom or from bottom to top, such a systematic approach may be cognitively easier for the DM for comparisons. In parallel to this, we start with comparing the topmost elements and go down in the hierarchy.

The topmost comparable elements in the hierarchy are "Product Related" and "User Related" criteria. Suppose the DM has provided the decision analyst with the pairwise comparison matrix in Table 7.

Table 7 Level-2 Pairwise comparison matrix

Mobile Phone Selection	**Product Related Criteria**	**User Related Criteria**
Product Related Criteria	1	$\frac{1}{2}$
User Related Criteria	2	1

MNV method starts with summing the numerical scores in each column of the pairwise comparison matrix. Respective column sums are noted under each row, as shown in Table 8.

Table 8 Level-2 Pairwise comparison matrix - Column summation

Mobile Phone Selection	**Product Related Criteria**	**User Related Criteria**
Product Related Criteria	1	$\frac{1}{2}$
User Related Criteria	2	1
Column Sums	**3**	$\boldsymbol{\frac{3}{2}}$

Then, each column is normalized to sum up to "1" by dividing each element by the respective column sum, calculated in Table 8. Normalized pairwise comparison matrix is generated as shown in Table 9.

Table 9 Level-2 Pairwise comparison matrix - Column normalization

Mobile Phone Selection	**Product Related Criteria**	**User Related Criteria**
Product Related Criteria	$\frac{1}{3}$	$\frac{1}{3}$
User Related Criteria	$\frac{2}{3}$	$\frac{2}{3}$
Column Sums	**1**	**1**

The last step to calculate weights is to average the numerical scores in each row. This step is simply carried out by summing all normalized numerical scores for each row and dividing the sum by the number of columns, as shown in Table 10.

Table 10 Level-2 Pairwise comparison matrix - Row averaging

Mobile Phone Selection	**Product Related Criteria**	**User Related Criteria**	**Row Means**
Product Related Criteria	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{\left(\frac{1}{3}+\frac{1}{3}\right)}{2}=\frac{1}{3}$
User Related Criteria	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{\left(\frac{2}{3}+\frac{2}{3}\right)}{2}=\frac{2}{3}$

Based on the calculations in Table 10, it can be concluded that "Product Related Criteria" and "User Related Criteria" have weights of approximately 0.333 and 0.666, respectively. This concludes the weight calculation step for the level-2 elements in the hierarchy. The next step is to calculate the weights in the remaining hierarchies. For the sake of simplicity, weights are calculated for the remaining pairwise comparisons and only the final weights are shown in Table 11.

Table 11 Calculated weights for level-3 of the hierarchy

Product Related Criteria	Weights	User Related Criteria	Weights
Basic Requirements	0.669	**Functionality**	0.633
Physical Characteristics	0.088	**Brand Choice**	0.260
Technical Features	0.243	**Customer Excitement**	0.106

So far, comparative weights of elements at each hierarchy level are calculated based on the criteria in their immediately upper hierarchy, named as the "parent criterion" (Schmidt et al., 2015). Yet, these weights are calculated only locally, as they represent how much they affect their parent criterion. Therefore, these weights are named as "local weights". Contribution of each criterion to the main goal of "Mobile Phone Selection" is still unknown.

As described by Saaty (1987), weights (priorities) are synthesized from level-2 down by multiplying local priorities by the priority of their corresponding parent criterion, for each element in a level according to the criteria it affects. This multiplication yields the overall contribution (global weight) of that criterion. Recall Figure 5. Previously calculated local weights of each criterion are shown in Figure 7. Note that the local weight of "Mobile Phone Selection" is "1.000" as it is the main goal itself.

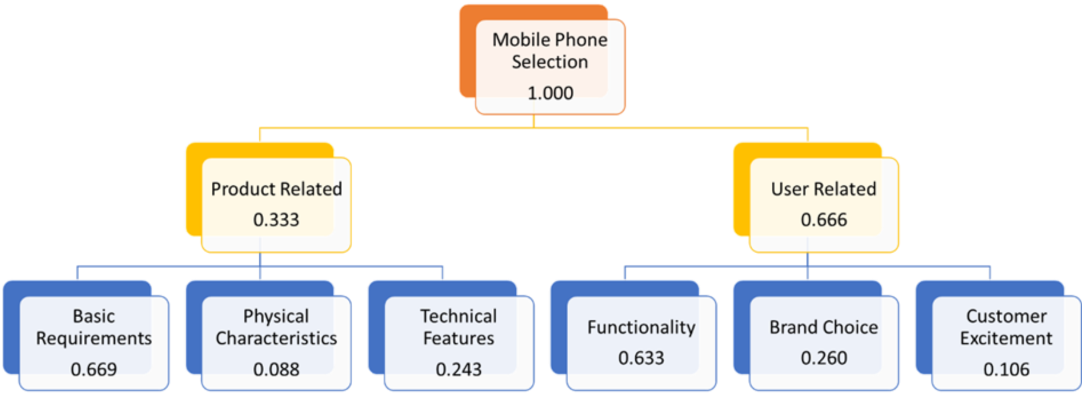

Figure 7 Local weights of each criterion in the hierarchy

Based on Figure 7, the global weight of "Basic Requirements" is calculated as follows:

$$w_{Basic\ Requirements} = 0.669 * \ 0.333 = 0.223$$

Carrying out the same operation for all lowest level criteria yields the global criteria weights in Table 12.

Table 12 Global criteria weights

Product Related Criteria	**Weights**	**User Related Criteria**	**Weights**
Basic Requirements	0.223	**Functionality**	0.422
Physical Characteristics	0.029	**Brand Choice**	0.173
Technical Features	0.081	**Customer Excitement**	0.072

As the lowest-level criteria's contribution to the main goal (global weights) are calculated, the only remaining operation for the decision is obtaining the overall score of each alternative. Similar to the previous application, all alternatives are compared in pairs with respect to each lowest-level sub-criteria. That is, Phone1, Phone2, and Phone3 are compared on their performances in "Basic Requirements", "Physical Characteristics", "Technical Features", "Functionality", "Brand Choice", and "Customer Excitement". These comparisons generate 6 more pairwise comparison matrices, and thus, 6 sets of local weights (scores) for 3 alternatives. Then, the local weights (scores) of each alternative is multiplied with its respective global weight and summed up to obtain the final score of the alternative. Suppose 6 pairwise comparisons resulted in the respective performances of the alternatives in each criterion as shown in Table 13.

Table 13 Local weights of alternatives in each lowest-level sub-criterion

Mobile Phone Selection	Criterion Weight	Phone1 Score	Phone2 Score	Phone3 Score
Basic Requirements	0.223	0.581	0.309	0.110
Physical Characteristics	0.029	0.315	0.602	0.082
Technical Features	0.081	0.089	0.324	0.587
Functionality	0.422	0.557	0.123	0.320
Brand Choice	0.173	0.416	0.126	0.458
Customer Excitement	0.072	0.174	0.723	0.103
Total	**1.000**	**0.465**	**0.238**	**0.296**

These results show that Phone1 is better than Phone2 and Phone3, considering the given pairwise comparisons. Although AHP suggests the best alternative based on the calculations, it is still up to the DM to make the decision.

As previously mentioned, EVM is the most widely used local weight determination method. Yet, the choice of weight determination method has been the topic of many debates, which contributed various approaches from different researchers. A study notes that (Srdjevic, 2005) EVM is the most common means of calculating the weights from PCMs, while some other methods (additive normalization, weighted least-squares, logarithmic least-squares, logarithmic goal programming, and fuzzy preference programming methods) also yield comparable results.

El Hafnawy and Mohammed (2014) classify weight derivation methods, mainly in two categories as shown in Figure 8. The study of El Hafnawy and Mohammed (2014) investigates the weight derivation methods, which are used at least in one AHP application.

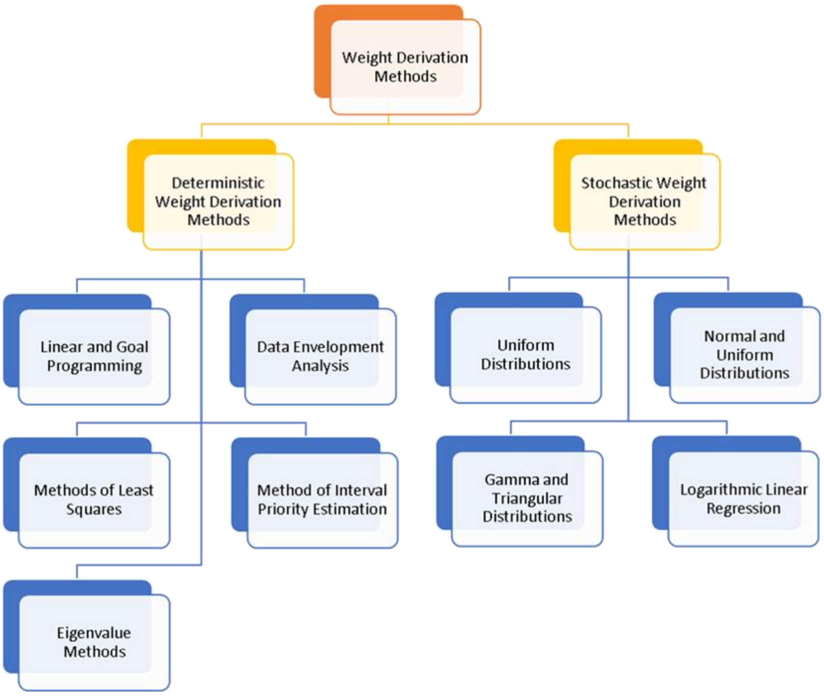

Figure 8 General classification of weight derivation methods

Taking a closer look in the wide AHP literature, on the other hand, it is observed that majority of the studies involve the derivation methods grouped under "Eigenvalue Methods" and "Methods of Least Squares". Based on the study of Ishizaka and Lusti (2006), these two groups of methods are further divided as shown in Figure 9.

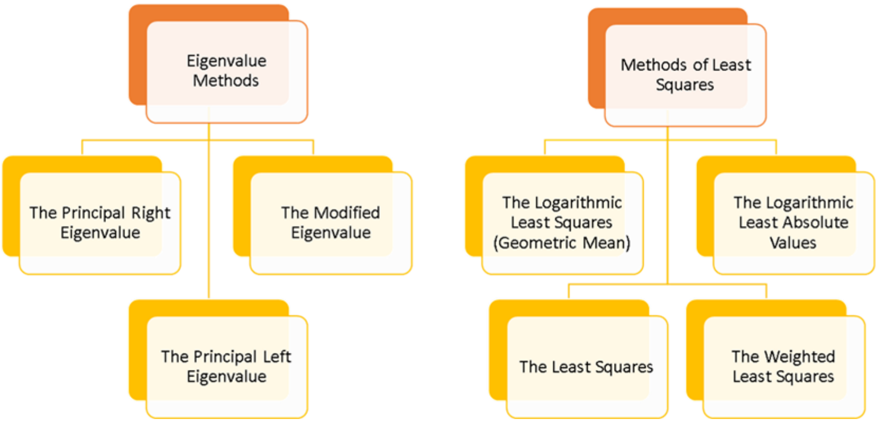

Figure 9 Classification of the frequently used weight derivation methods

Among these methods, it is noted that many studies focus on comparing EVM and LLSM. In a comparative study (Dong et al., 2008), it is stated that EVM and LLSM are the most commonly used weight derivation method. EVM is supported by several studies (Saaty, 1990, 2003, 2005; Kumar and Ganesh, 1996; Saaty and Hu, 1998), while LLSM is regarded as a better method by several other studies (Crawford and Williams, 1985; Takeda et al., 1987; Zahedi, 1986; Barzilai and Golany, 1997). Additionally, Herman and Koczkodaj (1995) emphasize that there is only a small difference between EVM and LLSM.

EVM and LLSM are described in the following sections.

2.3.3.1. Eigenvalue Method (EVM)

The Eigenvalue Method (EVM) is proposed by Saaty (1977; 1980) to derive local weights in AHP. Based on the EVM, weights are derived from pairwise comparison matrices based on the maximum principal (right) eigenvalue of a pairwise comparison matrix. Suppose matrix $\vec{A}$ is an $(n\,x\,n)$ pairwise comparison matrix, and $\vec{p}$ is its $(n\,x\,1)$ eigenvector corresponding to the maximum right eigenvalue of matrix $\vec{A}$. Matrix algebra dictates that the following equation is valid, where "m" is a real number:

$$\vec{A} * \vec{p} = m * \vec{p}$$

After the weight vector $\vec{p}$ is calculated, the elements are normalized by dividing each element by the sum of all elements of vector $\vec{p}$. This operation makes sure that the weights sum up to 1. Saaty claims that the corresponding weights are the normalized elements in the vector $\vec{p}$. He justifies the method based on the perturbation theory, according to which, slight changes in pairwise comparison matrices result in slight changes in the eigenvalues and corresponding eigenvectors.

Two years after the publication of Saaty's study (1977), Johnson, Beine and Wang (Johnson et al., 1979) criticized the EVM method as the right and left eigenvectors do not necessarily give the same result. This criticism lead to the suggestion of numerous weight derivation methods as an alternative to the EVM.

Despite its widespread use in real-life applications, Saaty's original approach of Eigenvalue Method (EVM) has some drawbacks and these drawbacks have been discussed by many authors in numerous studies. Some of these studies suggest partial solutions to what they criticized. Still, it is strongly noted by El Hefnawy and Mohammed (2014) that the EVM is the most used method by majority of applications.

2.3.3.2. Logarithmic Least Squares Method (LLSM)

The Logarithmic Least Squares Method (LLSM) is also known as Row Geometric Mean Method (RGMM). This approach is relatively simpler than the EVM. Vector of weights $\vec{w}$ is derived by obtaining row or column geometric means of the elements, where a_{ij} is the element on the ith row and jth column of the $(n\,x\,n)$ pairwise comparison matrix $\vec{A}$. Weight derivation using the LLSM is as follows:

$$\vec{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \rightarrow \vec{w} = \begin{bmatrix} \sqrt[n]{\prod_{j=1}^{n} a_{1j}} \\ \vdots \\ \sqrt[n]{\prod_{j=1}^{n} a_{nj}} \end{bmatrix}$$

After the weight vector $\vec{w}$ is calculated, the elements are normalized by dividing each element by the sum of all elements of vector $\vec{w}$.. This operation makes sure that the weights sum up to 1, as in the EVM. Although the method takes the row geometric means, column geometric means give the same result. Therefore, LLSM is regarded "insensitive" to the inversion of scale (Ishizaka and Lusti, 2006).

It is noted in several studies (Aguaron and Moreno-Jiménez, 2003; Alonso and Lamata, 2006) that there has been a significant increase in the use of LLSM as a weight derivation method in AHP. This increase can be attributed to its psychological and mathematical attributes (Aguaron and Moreno-Jiménez, 2003) as well as being more easily applicable compared to the other methods (Ishizaka and Lusti, 2006). An interesting fact is that the maximum left eigenvector of matrix $\vec{A}$ is approximated by taking the geometric mean of each row (Triantaphyllou and Mann, 1995).

Despite the heated debate on which method is better, there is still no agreement on a single weight derivation method for pairwise comparison matrices. According to researchers, each method has its advantages and disadvantages (El Hefnawy and Mohammed, 2014), and there is no superior method to another (Ishizaka and Lusti, 2006). Therefore, it is suggested that the weight derivation method should be chosen based on the application and other criteria like “easiness of use” should be considered during the method selection process.

2.3.4. Consistency Measurement

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 of the corresponding two elements, i and j. That is, for each numerical pairwise comparison score a_{ij} in the pairwise comparison matrix equals to

$$a_{ij} = \frac{w_i}{w_j}$$

where w_i and w_j are the final weights of the elements denoted on the ith and the jth rows of the pairwise comparison matrix, respectively.

By its nature, AHP is a human centered decision analysis tool. AHP requires subjective judgements, which takes into account the personal tastes, needs, experience, perception, specific knowledge and even the temporary mood of the DM. Considering the fact that the human mind is not a flawless measurement device and the abovementioned subjective items may vary with time, checking consistency is very important for the validity of the judgements. (El Hefnawy and Mohammed, 2014). As emphasized in Sagir Ozdemir's study (2005), although consistency of judgements may not be sufficient alone for the validity, a valid set of judgements must be consistent.

For a consistent matrix, it is obvious that only a complete row (or column) of pairwise comparisons is enough, since the other scores can be simply deduced from the ratios of available numerical comparison scores. Making all possible pairwise comparisons, on the other hand, brings redundancy in the information provided. Still, redundancy is regarded as a necessity to improve the validity of the outcome, particularly in the cases, which involve intangibles (Sagir Ozdemir, 2005).

By its definition, consistency in AHP requires complete transitivity in the pairwise comparison matrix. Saaty and Hu (1998) mention two kinds of transitivity:

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

Based on the consistency definition of AHP, a pairwise comparison matrix must be cardinally transitive, and therefore ordinally transitive, to be consistent. According to Saaty (2003), people are more likely to be cardinally inconsistent than cardinally consistent, as they cannot estimate precisely measurement values even from a known scale and worse when they deal with intangibles. As DMs are rarely fully consistent in their judgements, Saaty (1980) proposed the consistency index (CI) and the consistency ratio (CR) as a measure of the consistency of judgements. CI is defined by Saaty as:

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

where n is the dimension of $(n \, x \, n)$ pairwise comparison matrix and λ_{max} is the largest eigenvalue of this particular pairwise comparison matrix. Saaty and Vargas (2012) note that "$\lambda_{max} = n$" when the pairwise comparison matrix is consistent, i.e. all elements in the matrix are cardinally transitive. Under "$\lambda_{max} = n$" condition, the numerator of the CI equation becomes 0, which means perfect consistency. It is stated in the study of Alonso and Lamata (2006) that small changes in a numerical pairwise comparison judgement a_{ij} imply small changes in λ_{max}, which makes the difference between λ_{max} and n a good measure of consistency.

In addition to CI, Saaty suggested a normalization to the consistency measurement process, as the difference between λ_{max} and n tends to increase with the increasing matrix dimension. In order to normalize CI, Saaty suggested the CR, which measures the uniformity of a DM's answers as follows:

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

where RI is the random index, i.e. the average CI of the randomly filled PCMs. The main idea is that the CR is a normalized value since it is divided by an arithmetic mean of random matrix CIs (Alonso and Lamata, 2006). This way, Saaty enabled the normalization of the inconsistency of a pairwise comparison matrix with respect to an average inconsistency obtained from the random matrices of the same dimension.

Alonso and Lamata (2006) studied random index topic in detail and prepared a table of the RI values obtained in various main studies. The procedure is quite simple and consists of the following three steps:

- Random matrix generation (Saaty's Fundamental Scale, uniform distribution)
- Calculation of corresponding CI for each matrix
- Calculation of the mean of CI values for each matrix size

The number of randomly generated matrices and the RI values obtained for each matrix dimension are shown in Table 14.

Table 14 RI(n) values from various authors

	Random Matrices	Matrix Dimension (n)												
		3	4	5	6	7	8	9	10	11	12	13	14	15
Uppuluri (1980)	100	0.382	0.946	1.220	1.032	1.468	1.402	1.350	1.464	1.576	1.476	1.564	1.568	1.586
Saaty (1980)	500	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51				
Golden, Wang (1990)	100	0.5799	0.8921	1.1159	1.2358	1.3322	1.3952	1.4537	1.4882	1.5117	1.5356	1.5571	1.5714	1.5831
Lane, Verdini (1989)	2,500	0.52	0.87	1.10	1.25	1.34	1.40	1.45	1.49		1.54		1.57	
Forman (1990)		0.5233	0.8860	1.1098	1.2539	1.3451								
Noble, Sanchez (1993)	500	0.49	0.82	1.03	1.16	1.25	1.31	1.36	1.39	1.42	1.44	1.46	1.48	1.49
Tumala, Wan (1994)		0.500	0.834	1.046	1.178	1.267	1.326	1.369	1.406	1.433	1.456	1.474	1.491	1.501
Aguaron et al (2003)	100,000	0.525	0.882	1.115	1.252	1.341	1.404	1.452	1.484	1.513	1.535	1.555	1.570	1.583
Alonso, Lamata (2004)	100,000	0.5245	0.8815	1.1086	1.2479	1.3417	1.4056	1.4499	1.4854	1.5141	1.5365	1.5551	1.5713	1.5838
Alonso, Lamata (2006)	500,000	0.5247	0.8816	1.1086	1.2479	1.3417	1.4057	1.4499	1.4854	1.5140	1.5365	1.5551	1.5713	1.5838
Franek, Kresta (2014)	500,000	0.525	0.882	1.110	1.250	1.341	1.404	1.451	1.486	1.514	1.536	1.555	1.570	1.584

Sources: Alonso and Lamata (2006); Franek and Kresta (2014)

Based on the RI values calculated by various authors, it can be concluded that the values vary between different experiments. However, at higher numbers of randomly generated matrices, it is observed that RI values converge to some certain values. Additionally, it should be noted that these values are only valid for Saaty's Fundamental Scale. Therefore, if any other scale is to be used, RI values should be calculated in the same way. Salo and Hämäläinen (1997) emphasize that the CR is a meaningful measure only if the same scale has been employed both in the assessment of the actual comparison matrix and in the generation of the random matrices to calculate RI.

In Saaty's consistency measure based on the eigenvalue approach, the closer CR to 0 the more consistent the judgements of the DM. However, as people are seldom fully consistent in their judgements, a cut-off point is required to separate consistent and inconsistent pairwise comparison matrices, such that beyond this particular cut-off point the pairwise comparison matrix is regarded as inconsistent. The most commonly used cut-off point is the one proposed by Saaty (1980). According to Saaty, small values of inconsistency may be tolerated. Particularly, if CR is less than 10% (0.1), inconsistencies are tolerable. If the CR exceeds this threshold, then the matrix is deemed unacceptably inconsistent. As consistency is necessary for a pairwise comparison matrix to be valid, in case of an unacceptable inconsistency, the DM is advised to revise the pairwise comparison matrix. This step is repeated until the DM provides an acceptably consistent matrix or is certain that no more revisions can be made. Later in another study, Saaty (1996) proposed 5% and 8% as thresholds for 3x3 and 4x4 matrices respectively, while keeping it 10% for larger matrices.

According to Dodd et al. (1992), this 10% threshold was clearly intended to be only a tentative measure. Yet, it is surprising that this tentative measure has been widely accepted without much questioning. In the following years, several other methods have been proposed for consistency measurement. One of the most well-known approaches is the geometric consistency index (GCI) proposed by Crawford and Williams (1985), who preferred to sum the difference between the ratio of calculated weights and the pairwise comparison matrix provided by the DM. GCI proposed by Crawford and Williams (1985) is formulated as:

$$GCI = \frac{2\sum_{i<j}\left(log\ a_{ij} - log\frac{w_i}{w_j}\right)^2}{(n-1)(n-2)}$$

Based on this formulation of GCI, Aguarón and Moreno-Jiménez (2003) calculated geometric thresholds which are analogous to CR=10%. They calculated GCI values as:

- GCI=0.3147 for *n*=3
- GCI=0.3526 for *n*=4
- GCI=0.3700 for *n*>4

Another formulation of GCI has been proposed by other studies (Salo and Hämäläinen, 1997; Ji and Jiang, 2003), which is moderately different than the previous GCI formula:

$$GCI = \sqrt{\frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left(a_{ij} - log\frac{p_i}{p_j}\right)}{\frac{n(n-1)}{2}}}$$

In another study (Dodd et al., 1992), a statistical approach to consistency has been suggested as an alternative to this tentative threshold. This statistical approach employs confidence intervals to determine if a matrix is consistent, based on normal distribution of the CI values of randomly generated pairwise comparison matrices. The CI values of the pairwise comparison matrix is compared to the tolerance threshold (95%, 99% or 99.9%) selected by the DM. Then the pairwise comparison matrix is accepted if it has a lower CI than required, and rejected otherwise.

An important and influential question is raised by the literature: Why tolerate 10% inconsistency? Can there be another cut-off point based on the accuracy need of the problem at hand? This question is partially answered by a different approach (Alonso and Lamata, 2006), which proposed a formulation for λ_{max} with an adaptive consistency coefficient α that is determined by the DM based on the level of consistency needed. According to this formulation, matrices ensuring the following condition are acceptably consistent:

$$\lambda_{max} \leq n + \alpha(1.7699n - 4.3513)$$

Other methods can be listed as:

- Consistency check based on the determinant of the pairwise comparison matrix (Peláez and Lamata, 2003)
- Accepting/rejecting the pairwise comparison matrix based on the average principal eigenvalues (Alonso and Lamata, 2006)
- Harmonic consistency index (HCI) based on additive normalization method (Stein and Mizzi, 2007)

It can be deduced from the literature review that the consistency and accepting/rejecting matrices topics are highly debated in AHP literature. Lane and Verdini (1989) state that Saaty's 10% threshold is too restrictive as the standard deviation of the CI values of randomly generated matrices are small. Another study (Murphy, 1993) shows that Saaty's Fundamental Scale gives results which are beyond the CR acceptance threshold as matrix dimension "n" increases. Salo and Hämäläinen (1997), on the other hand, draw attention to the fact that the CR acceptance threshold depends on the granularity of the scale used for pairwise comparisons. Another surprising result obtained in a recent study (Mazurek and Perzina, 2017) is that DMs are mostly inconsistent (93%) even when they are only asked to make three pairwise comparisons. As such, the authors conclude with a remark implying that cardinal transitivity may be too strong for practical use, and it might need to be substituted by ordinal transitivity.

Still, Saaty (2003) emphasizes that "if one insists on consistency, people would be required to be like robots unable to change their minds with new evidence and unable to look within for judgements that represent their thoughts, feelings and preferences".

Therefore, a slight amount of inconsistency may be considered a good thing and forced consistency may be an undesirable compulsion for the validity of a pairwise comparison matrix.

2.3.5. Aggregation of the Local Priorities

The last step in AHP is to aggregate (synthesize) the local priorities (weights) across all criteria to calculate the overall scores of alternatives with respect to the goal of the decision making process. Similar to the weight derivation methods, how to aggregate the overall score of an alternative has been widely discussed in the literature. Basically, there are two main approaches to obtain the overall scores of alternatives from pairwise comparison matrices, which can be named as “Additive Aggregation” and “Multiplicative Aggregation” (Choo and Wedley, 2008).

2.3.5.1. Additive Aggregation

Additive aggregation is the method that was brought to the literature by the original AHP approach. Additive aggregation employs a weighted additive method to aggregate the overall scores. Recall Table 13, where the global weights of all sub-criteria and the respective local weights of Phone1 in each sub-criterion are shown as in Table 15.

Table 15 Global sub-criteria weights and Phone1’s respective performances on each sub-criterion

Sub-Criterion	Criterion Weight	Phone1 Score
Basic Requirements	0.223	0.581
Physical Characteristics	0.029	0.315
Technical Features	0.081	0.089
Functionality	0.422	0.557
Brand Choice	0.173	0.416
Customer Excitement	0.072	0.174

Based on additive aggregation approach, the overall score of Phone1 is calculated by multiplying its score in each sub-criterion with the global weight of the respective sub-criterion. That yields the overall score of Phone1 as follows:

$$(0.223 * 0.581) + (0.029 * 0.315) + (0.081 * 0.089) + (0.422 * 0.557) + (0.173 * 0.416) + (0.072 * 0.174) = 0.465$$

Additive aggregation approach has been widely criticized for its famous "rank reversal phenomenon". It is stated in the study of Ishizaka et al. (2010) that the rank reversal problem in AHP is argued as it is due to incorrect usage of the additive aggregation method. According to Tomashevskii (2015) all EVM rank reversal phenomena have the same cause. Another study (Tomashevskii, 2014) further divides the rank reversal problem into three different categories as follows:

- Rank reversal for scale inversion or the right-left eigenvector asymmetry (Johnson et al., 1979)
- Rank reversal caused by the addition or deletion of an element under consideration (Hochbaum and Levin, 2006; Raharjo and Endah, 2005)
- Rank reversal of "order of intensity of preference" (Bana e Costa and Vansnick, 2008)

2.3.5.2. Multiplicative Aggregation

Multiplicative aggregation has been proposed as an alternative to additive aggregation, with the main motivation of preventing the rank reversal. According to Ishizaka and Labib (2011), multiplicative aggregation approach has non-linearity properties, which allow a superior compromise to be selected.

Multiplicative aggregation approach calculates the overall score of Phone1 by multiplying its score in each sub-criterion using the global weight of the respective sub-criterion as its exponent. Using the same example given in Table 15, the overall score of Phone1 is calculated by multiplicative aggregation approach as follows:

$$(0.581)^{0.223} * (0.315)^{0.029} * (0.089)^{0.081} * (0.557)^{0.422} * (0.416)^{0.173} * (0.174)^{0.072} = 0.417$$

Similarly, scores of Phone2 and Phone3 are calculated as 0.195 and 0.250, respectively. Dividing each final score to the sum of the scores of all three alternatives, normalized scores appear as shown in Table 16.

Table 16 Aggregated scores of all alternatives by additive and multiplicative aggregation approaches

Aggregation Method	**Phone1 Score**	**Phone2 Score**	**Phone3 Score**
Additive Aggregation	0.465	0.238	0.296
Multiplicative Aggregation	0.484	0.226	0.250

Apparently, for our example, aggregation method does not change the rank of any alternatives. Yet, it should be noted that the differences between the weights of alternatives are less when additive aggregation is employed as the aggregation method.

CHAPTER 3

LIMITATIONS OF EXISTING SCALES

According to Stevens (1957) and Stevens and Galanther (1964), a ratio scale is an appropriate means that can be used to elicit response stimuli. The pairwise comparison scale in AHP is defined as a ratio scale, and it is assumed that one can express the cardinal intensity of preference between two compared elements by using this ratio scale. Similarly, Ji and Jiang (2003) emphasize that the values defined on a pairwise comparison scale are used to represent the relative importance between two objects in terms of a ratio. In parallel to this statement, Franek and Kresta (2014) note that AHP requires a ratio scale due to its pairwise comparison characteristic.

The reason why a 1-9 scale (Fundamental Scale) is chosen has been explained in the study of Harker and Vargas (1987). They point out that although Saaty chose Fundamental Scale based on experimental evidence, any ratio scale can be used in AHP. Another point they emphasized is that any bounded ratio scale would be in accordance with the axioms of AHP.

Saaty supports his Fundamental Scale with famous Weber-Fechner Psychophysical Law. According to Weber's approach, an observer's ability to detect a difference between two different stimuli is a function of an observer-specific constant "k" and the value of the standard stimulus "S". Mathematically, Weber defines the "just noticeable difference" (*JND*) as:

$$JND = k * S$$

According to Weber's formula, if the difference between 100 and 105 grams is just noticeable by an observer, the same observer can notice the difference between 1000 and 1050 grams. In this example, if the difference between two stimuli is less than 5%

of the value of standard stimulus, then the observer would not be able to notice the difference.

Later, Fechner (1860) derived a relationship between the intensity of a stimulus and its perceived magnitude by assuming that Weber's Law holds and the JND is the basic unit of perceived magnitude so that one JND is perceptually equal to another JND. Fechner defines the mathematical relationship between the perceived magnitude of a stimuli "P" and the stimulus intensity "I" as:

$$P = k * \log(I)$$

Based on Fechner's formulation, the perceived magnitude and the stimulus intensity have a diminishing relationship. For instance, if the intensity of light is doubled in an environment, let's say from 15 to 30, the perceived magnitude of this increase would not be directly proportional. As such, the effect would be:

$$\frac{P_2}{P_1} = \frac{k * \log(30)}{k * \log(15)} = 1.256$$

According to Saaty's approach based on Stevens (1957) and Fechner (1860), if a stimulus is increased successively from one point to the next "detectable difference" point, then the points on the stimulus scale are geometrically related and the points on the response scale are linear. In AHP methodology, "P" is analogous to the answers of the DM to pairwise comparison questions, and "I" is analogous to the real value of the stimulus. Saaty's approach would be possible only if the "I" values of two different stimuli are on the same geometric function, such as 10^x for our example. Then, what the DM perceives from the difference between two different stimuli and what answer he/she gives would obey an arithmetic progression.

Although it is easy to understand and use in practical situations, Fundamental Scale has another major flaw. Dodd et al. (1992) criticize Saaty for his overestimation of the validity of normal arithmetic on his set. As emphasized, the scale used in AHP methodology is a ratio scale. However, ratios between the successive elements of Fundamental Scale are not the same. Ji and Jiang (2003) draws attention to the fact that the difference of the importance (or preference) intensities of any two adjacent major gradations in Fundamental Scale (1, 3, 5, 7, 9) in the verbal part form an arithmetic progression. They call this “the arithmetic progression rule of the verbal part”. It is indeed the case in the numerical part as well. From 1 to 9, the numerical responses form an arithmetic progression as well. However, considering the fact that AHP methodology employs reciprocals of the numerical scores to complete the pairwise comparison matrix, the numerical part below 1 (i.e. 1/2, 1/3, … , 1/9) does not form the same arithmetic progression. Dodd et al. (1992) criticize Fundamental Scale as it is partially linear and partially harmonic, which in turn, already disturbs the “ratio” nature of pairwise comparisons in AHP. This characteristic will be referred in this study as “the partial characteristic of Fundamental Scale”.

Another point that has been criticized by several authors (French, 1988; Dodd et al., 1992; Ji and Jiang, 2003) is that Fundamental Scale is not closed under multiplication. This problem is also named as “the boundary problem”. That is, Fundamental Scale has an inherent tendency to inconsistency as the scale has upper and lower limits. For instance, considering the cardinal transitivity requirement for consistency, if “$a_{ij} = 4$” and “$a_{jk} = 5$”, then “$a_{ik} = 20$” must hold to enable consistency. However, Ji and Jiang (2003) state that all of the existing scales suffer from this type of inconsistency. Sagir Ozdemir (2005), on the other hand, notes that people are unable to directly compare widely disperate objects. In order to do that, she claims AHP needs a range greater than 1-9 scale (Fundamental Scale).

Table A3 in Appendix A section includes the possible combinations of (a_{ij}, a_{jk}) and the resultant a_{ik} values. Based on our analysis, it can be concluded that only 173 of possible 289 combinations are within the limits of Fundamental Scale. A significant portion (126 combinations), however, is beyond upper or lower limits, that result in a huge possibility of inherent inconsistency.

Another cause of inconsistency is due to the discreteness of the scale. That is, even if “$a_{ik} = a_{ij} * a_{jk}$” is within the upper and lower limits of the scale, “a_{ik}” may not be one of the numerical scores defined by the scale itself. They show that 44 of 81 possible multiplication results are not defined by Fundamental Scale, as shown in Table 17. They add that a geometric scale would not have such inconsistencies. That is to say, if a multiplication is within the limits of the scale, there is no chance that the multiplication is not defined by the scale as a numerical pairwise comparison ratio.

Table 17 Values of (a_{ij} * a_{jk}) in Fundamental Scale

a_{ik}	**1**	**2^{-1}**	**3^{-1}**	**4^{-1}**	**5^{-1}**	**6^{-1}**	**7^{-1}**	**8^{-1}**	**9^{-1}**
1	1	1/2	1/3	1/4	1/5	1/6	1/7	1/8	1/9
2	2	1	2/3	2/4	2/5	2/6	2/7	2/8	2/9
3	3	3/2	1	3/4	3/5	3/6	3/7	3/8	3/9
4	4	2	4/3	1	4/5	4/6	4/7	4/8	4/9
5	5	5/2	5/3	5/4	1	5/6	5/7	5/8	5/9
6	6	3	2	6/4	6/5	1	6/7	6/8	6/9
7	7	7/2	7/3	7/4	7/5	7/6	1	7/8	7/9
8	8	4	8/3	2	8/5	8/6	8/7	1	8/9
9	9	9/2	3	9/4	9/5	9/6	9/7	9/8	1

Source: Ji and Jiang (2003)

According to Budescu et al. (1986) and Crawford (1987), the scale has an inherent exponential quality. What they mean by the “exponential quality” is that a consensus in a group decision making can be found by geometrically averaging the experts’ judgements, and thus, the scale should be compatible with multiplicative/divisive operations.

Exponential pairwise comparison scale idea first appears in Lootsma (1989). The formulation he discusses is:

$$r_{ij} = e^{\lambda \delta_{ij}}$$

where δ_{ij} is an integer designating the gradation chosen by the DM to estimate the ratio difference between items $i - j$, λ is a scale constant, and r_{ij} is the numerical score of the ith row jth column element of the pairwise comparison matrix. According to Lootsma, if a trade-off estimate between two items is carried out by employing an exponential scale, the additive degree of freedom can clearly be ignored, and thus, the true ratio difference between these items can be estimated. That is:

$$t_{ij-km} = \frac{e^{\lambda\delta_{ij}}}{e^{\lambda\delta_{km}}} = e^{\lambda(\delta_{ij}-\delta_{km})}$$

where t_{ij-km} is the ratio between the values of elements ij and km. Note that the numerical scores of elements ij, km and the ratio of their numerical scores t_{ij-km} are all on an exponential function. This kind of a function eliminates the previously mentioned reservations about the partial characteristic of Fundamental Scale.

Although Lootsma (1989) notes that there is no unique value of the scale constant "λ", and expresses his feeling about λ=1 or λ=2 would be appropriate choices, this selection of the scale constant appear too "intuitive" as it finally affects the numerical score of the comparison in the pairwise comparison matrix. Therefore, there is room for an improvement for a method to determine the scale constant "λ".

Ji and Jiang (2003), on the other hand, note that an AHP scale can hold the transitivity if:

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

Therefore, a geometric/exponential progression is desirable for the numerical part of a pairwise comparison scale.

For the Geometric Scale of Lootsma (1989), the most important problem is that the upper limit of 64 or 256 severely violates the homogeneity axiom of AHP (Ji and Jiang, 2003). Another criticism about the Geometric Scale is that it is obtained from assumptions and external observations, thus, making its theoretical foundations rather weak.

CHAPTER 4

A NEW PAIRWISE COMPARISON SCALE BASED ON FIBONACCI SEQUENCE

The famous Fibonacci sequence 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:

$$F_n = F_{n-1} + F_{n-2}$$

where the first two numbers in the series are initially defined as $F_1 = 1$ and $F_2 = 1$. This makes the series appear as follows:

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

By nature, Fibonacci sequence has an exponential characteristic. In order to show this with a smooth exponential equation, we adjust the one-to-one mapping between the verbal pairwise comparison scales and numerical grades. We use a similar approach adopted as in Ji and Jiang (2003), that is to “digitize” the verbal scale by using a range of [-4, 4] with increments of 0.5. However, in our study, we use a range of [-8, 8] with increments of 1 in the digitized part. Table 18 shows the original and adjusted one-to-one mapping between the verbal scale and its digitized counterpart.

Table 18 Adjusted one-to-one mapping between the verbal scale and numerical grades

Verbal Scale	Numerical Grades (Original)	Digitized Verbal Part (Ji and Jiang, 2003)	Numerical Grades (Adjusted)
Equally important	1	0	0
Slightly more (less) important	3	1 (-1)	2 (-2)
Strongly more (less) important	5	2 (-2)	4 (-4)
Very strongly more (less) important	7	3 (-3)	6 (-6)
Absolutely more (less) important	9	4 (-4)	8 (-8)
Compromises	2, 4, 6, 8	±0.5 , ±1.5, ±2.5, ±3.5	±1 , ±3, ±5, ±7

Using the adjusted numerical grades shown in Table 18, a curve is fit to the first nine Fibonacci numbers (excluding the first "1") and their multiplicative inverses (reciprocals). Curve fitting operation is carried out by using MATLAB® software for the sake of accuracy. Figures 10 and 11 show the exponential curve fit to the Fibonacci numbers based on the adjusted numerical grades and the equation/goodness of fit data retrieved from MATLAB® Curve Fitting Tool screen.

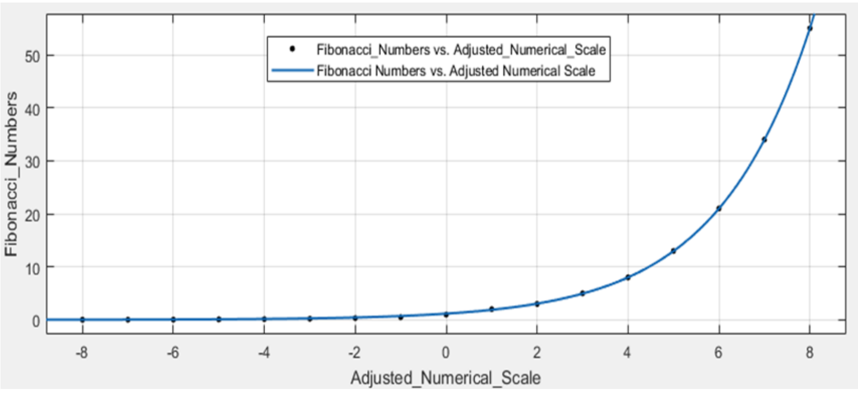

Figure 10 Exponential fit to the Fibonacci numbers based on the adjusted numerical grades

```
General model Exp1:
   f(x) = a*exp(b*x)
Coefficients (with 95% confidence bounds):
     a =       1.166  (1.142, 1.189)
     b =      0.4818  (0.4791, 0.4845)

Goodness of fit:
  SSE: 0.1177
  R-square: 1
  Adjusted R-square: 1
  RMSE: 0.08857
```

Figure 11 Equation and goodness of exponential fit

As seen in Figure 10, all of the numerical values are on the same function, unlike Fundamental Scale. Thus, we have already overcome the partial characteristic issue criticized by Dodd et al. (1992).

Evaluating the exponential curve fit to the Fibonacci numbers using the adjusted numerical value grades, the initial approximate numerical values to be used in pairwise comparisons (a_{ij}) are obtained as shown in Table 19.

Table 19 Approximate numerical pairwise comparison values (Initial)

Verbal Scale	**Adjusted Numerical Grade**	**Approximate Numerical Value (a_{ij})**
Absolutely less important	-8	0.025
	-7	0.040
Very strongly less important	-6	0.065
	-5	0.105
Strongly less important	-4	0.170
	-3	0.275
Slightly less important	-2	0.445
	-1	0.720
Equally important	0	1.166
	1	1.888
Slightly more important	2	3.056
	3	4.948
Strongly more important	4	8.011
	5	12.969
Very strongly more important	6	20.997
	7	33.994
Absolutely more important	8	55.036

Due to the coefficient “1.166” in the exponential equation, it is seen that the approximate numerical value of a_{ij} for equally important alternatives is “1.166” instead of “1.000”. Similarly, for all other possible a_{ij}, values are multiplied with “1.166”. Therefore, an additional adjustment is necessary. As such, dividing the fit exponential curve by the coefficient “1.166”, and thus dividing all elements on the curve, we obtain the following equation:

$$a_{ij} = \frac{1.166e^{0.4818\delta_{ij}}}{1.166} = e^{0.4818\delta_{ij}}$$

where a_{ij} for the equally important situation, naturally, becomes equal to “1.000”.

With this final adjustment, the equation becomes principally the same with what was discussed by Lootsma (1989):

$$r_{ij} = e^{\lambda\delta_{ij}} \longrightarrow a_{ij} = e^{0.4818\delta_{ij}}$$

where δ_{ij} is the adjusted numerical grade value that is mapped to the verbal answer of the DM. Therefore, as mentioned previously, instead of selecting $\lambda = 1$ or $\lambda = 2$ intuitively, we have now determined the scale constant λ based on the Fibonacci sequence. Thus, the final approximate numerical values to be used in pairwise comparisons (a_{ij}) become as shown in Table 20. This exponential scale based on the numbers of Fibonacci sequence is referred to as “Exponential Scale” in the rest of this study.

Table 20 Approximate numerical pairwise comparison values (Final)

Verbal Scale	Adjusted Numerical Grade	Approximate Numerical Value (a_{ij})
Absolutely 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
Absolutely more important	8	47.200

4.1. Scale-Based Inconsistency in the Existing Scales and Exponential Scale

In our previous discussion, the inconsistency characteristic due to the boundary problem of the existing scales were discussed. Recall it was mentioned that all of the existing scales suffer from this type of inconsistency (Ji and Jiang, 2003). Therefore, the main aim at this point should be to reduce, if elimination is not possible, the inconsistency based on the boundary issue. In order to show how much the existing scales and the Exponential Scale are affected by this issue, an analysis is made on all numerical AHP scales.

All scales are evaluated on boundary issue by checking how many of all paired-combination multiplications of scale values remain within the scale boundaries. That is, for each possible (a_{ij}, a_{jk}) paired-combination, it is checked whether their multiplication "$a_{ik} = a_{ij}\ x\ a_{jk}$" remains within the upper and lower boundaries of the same scale. For instance, $a_{ik} = 8$ where $a_{ij} = 4$ and $a_{jk} = 2$ is regarded as "within the limits" while $a_{ik} = 15$ where $a_{ij} = 3$ and $a_{jk} = 5$ is regarded as "outside the limits". Based on the analysis, the results in Table 21 are obtained:

Table 21 Possible paired-combination multiplication results within limits of the scale

Scale	**Number of possible combinations**	**Number of multiplications within limits**	**Percent multiplications within limits**	**Analysis Results**
Balanced	289	231	79.93%	Table A 1
Exponential	289	217	75.09%	Table A 2
Fundamental	289	173	59.86%	Table A 3
Geometric	289	217	75.09%	Table A 4
Inverse Linear	289	249	86.16%	Table A 5
Logarithmic	289	167	57.79%	Table A 6
Power	289	173	59.86%	Table A 7
Root Square	289	173	59.86%	Table A 8

Results clearly indicate that Fundamental, Logarithmic, Power and Root Square scales are very susceptible to inconsistency due to the boundary problem. Inverse Linear Scale, on the other hand, has a surprisingly high percentage of paired-combination multiplication results within limits, making it superior to the other scales for this measure. Exponential and Geometric scales are very similar to each other in this characteristic, while Balanced Scale is slightly better than these scales. The analysis results can be seen from the respective tables in the Appendices section.

Another inconsistency reason is the discreteness of the scale, as named by Ji and Jiang (2003). According to them, even if the result of a multiplication "$a_{ik} = a_{ij}\ x\ a_{jk}$" is within the limits of the scale, the result not being one of the values defined by the scale causes an inherent inconsistency. For instance, if $a_{ij} = 4$ and $a_{jk} = 1/3$, then their multiplication "$a_{ik} = a_{ij}\ x\ a_{jk} = 1.333$". It is clear that the result is not a member of

Fundamental Scale values. Thus, the closest value of a_{ik} is either "1" or "2", which is already inconsistent for both values.

Recall the analysis of Ji and Jiang (2003) in Table 17, where they argued that 44 of 81 possible multiplication results are not defined by Fundamental Scale. In our study, we further elaborated on their analysis and checked the percentage of paired-combination multiplication results, which are within the limits of a scale and defined by the scale as a numerical comparison value. The analysis yield the results shown in Table 22:

Table 22 Possible paired-combination multiplication results defined by the scale

Scale	**Number of possible combinations**	**Number of multiplications defined by the scale**	**Percent multiplications defined by the scale**	**Analysis Results**
Balanced	289	55	19.03%	Table A 1
Exponential	289	217	75.09%	Table A 2
Fundamental	289	85	29.41%	Table A 3
Geometric	289	217	75.09%	Table A 4
Inverse Linear	289	73	25.26%	Table A 5
Logarithmic	289	61	21.11%	Table A 6
Power	289	85	29.41%	Table A 7
Root Square	289	85	29.41%	Table A 8

These results show that for all scales except the Exponential and Geometric scales are badly affected by inherent inconsistency due to the multiplication values not defined on the scale values. The Exponential and Geometric scales, on the other hand, are stable in scale-based inconsistency issues with all multiplication values within the scale being defined by the scale itself. Based on the results, it can be said that even Inverse Linear and Balanced Scales have inherent inconsistency characteristics. The analysis results can be seen from the respective tables in the Appendices section.

CHAPTER 5

SIMULATION RESULTS

In this chapter, the results of simulations regarding the consistency characteristics of different pairwise comparison scales are presented. Firstly, previous results for RI values and the RI values generated during this study are compared in order to make sure that numerically the same or similar RI values are used in further calculations. Then, performance measures are explained and simulation steps are clarified. All scales are evaluated based on the performance measures. Lastly, these evaluations are shown visually in graphs and the performance characteristics are elaborated on.

5.1. Comparison of All Scales with Fundamental Scale

Consistency is an important issue, however, it is not alone enough to claim that one scale is superior to another. Although Fundamental Scale has consistency issues, it is still the most preferred scale in AHP applications. Therefore, a complete comparison is necessary to support the idea that a scale is better than Fundamental Scale. Thus, in the following part of this study, Fundamental Scale is used as a benchmark to compare all scales.

In order to compare the scales with Fundamental Scales, a detailed simulation is carried out using MATLAB® software. The detailed results of these simulations can be seen in Appendix B.

The first task was to generate each scale using the respective formula given in Table 6 and the corresponding numerical values are provided in Table 23. Note that these values not only cover the "more important" cases but also the "less important" cases.

Table 23 Numerical values for each scale, generated by the simulation

Verbal Scale	Balanced	Exponential	Fundamental	Geometric	Inverse Linear	Logarithmic	Power	Root Square
Absolutely less important	0.111	0.021	0.111	0.004	0.111	0.301	0.012	0.333
	0.176	0.034	0.125	0.008	0.222	0.315	0.016	0.354
Very strongly less important	0.250	0.056	0.143	0.016	0.333	0.333	0.020	0.378
	0.333	0.090	0.167	0.031	0.444	0.356	0.028	0.408
Strongly less important	0.429	0.146	0.200	0.063	0.556	0.387	0.040	0.447
	0.538	0.236	0.250	0.125	0.667	0.431	0.063	0.500
Slightly less important	0.667	0.382	0.333	0.250	0.778	0.500	0.111	0.577
	0.818	0.618	0.500	0.500	0.889	0.631	0.250	0.707
Equally important	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.222	1.619	2.000	2.000	1.125	1.585	4.000	1.414
Slightly more important	1.500	2.621	3.000	4.000	1.286	2.000	9.000	1.732
	1.857	4.244	4.000	8.000	1.500	2.322	16.000	2.000
Strongly more important	2.333	6.870	5.000	16.000	1.800	2.585	25.000	2.236
	3.000	11.123	6.000	32.000	2.250	2.807	36.000	2.449
Very strongly more important	4.000	18.008	7.000	64.000	3.000	3.000	49.000	2.646
	5.667	29.154	8.000	128.000	4.500	3.170	64.000	2.828
Absolutely more important	9.000	47.200	9.000	256.000	9.000	3.322	81.000	3.000

The second step is the calculation of RI values for each scale. As stated before, RI calculation consists of three steps:

- Random matrix generation (Saaty's Fundamental Scale, uniform distribution)
- Calculation of corresponding CI for each matrix
- Calculation of the mean of CI values for each matrix size

Franek and Kresta (2014) estimated the RI values for all scales (except for the new proposed Exponential Scale) with their simulation as shown in Table 24.

Table 24 Previously calculated RI values for each scale

Matrix Dimension (n)	Balanced	Fundamental	Geometric	Inverse Linear	Logarithmic	Power	Root Square
3	0.267	0.525	4.592	0.203	0.153	3.609	0.114
4	0.440	0.881	9.299	0.333	0.241	6.987	0.179
5	0.550	1.110	13.322	0.417	0.295	9.464	0.218
6	0.625	1.250	16.500	0.475	0.328	11.049	0.243
7	0.676	1.341	18.897	0.517	0.351	12.071	0.260
8	0.715	1.404	20.714	0.547	0.368	12.748	0.273
9	0.743	1.451	22.089	0.572	0.380	13.221	0.282
10	0.765	1.486	23.152	0.590	0.390	13.567	0.290
11	0.783	1.514	23.958	0.605	0.398	13.833	0.296
12	0.797	1.536	24.607	0.617	0.405	14.039	0.301
13	0.810	1.555	25.117	0.627	0.410	14.211	0.305
14	0.820	1.570	25.539	0.636	0.415	14.346	0.309
15	0.829	1.584	25.871	0.643	0.419	14.461	0.312

Source: Franek and Kresta (2014)

In order to make sure that our RI calculation for the Exponential Scale, as well as the others, are in parallel to what was found by Franek and Kresta (2014), we ran a simulation including all scales, for which, RI values were previously calculated. The values in Table 25 show our results. The percentages in Table 25 indicate the percent deviation between the values found by Franek and Kresta (2014) and the values calculated in this study. These deviations are calculated in two steps:

- Dividing the values in Table 24 by the values in Table 25 and
- Taking the absolute value of the difference between unity and ratios

Table 25 RI values calculated based on our simulation

Matrix Dimension (n)	Balanced	Exponential	Fundamental	Geometric	Inverse Linear	Logarithmic	Power	Root Square
3	0.268 0.38%	1.399	0.524 0.28%	4.565 0.59%	0.206 1.32%	0.153 0.09%	3.614 0.13%	0.115 0.45%
4	0.440 0.00%	2.572	0.885 0.47%	9.295 0.04%	0.333 0.08%	0.242 0.33%	6.996 0.13%	0.179 0.27%
5	0.550 0.04%	3.431	1.109 0.13%	13.323 0.01%	0.418 0.15%	0.295 0.14%	9.449 0.16%	0.218 0.11%
6	0.624 0.08%	4.042	1.249 0.11%	16.536 0.22%	0.475 0.03%	0.328 0.08%	11.050 0.01%	0.243 0.07%
7	0.676 0.07%	4.470	1.341 0.03%	18.918 0.11%	0.516 0.17%	0.351 0.04%	12.076 0.04%	0.260 0.08%
8	0.714 0.11%	4.777	1.404 0.02%	20.688 0.12%	0.547 0.07%	0.368 0.06%	12.752 0.03%	0.273 0.13%
9	0.743 0.01%	5.008	1.451 0.03%	22.107 0.08%	0.571 0.13%	0.381 0.15%	13.220 0.01%	0.282 0.04%
10	0.765 0.01%	5.174	1.486 0.02%	23.158 0.03%	0.590 0.04%	0.390 0.08%	13.569 0.01%	0.290 0.09%
11	0.783 0.01%	5.313	1.514 0.01%	23.964 0.02%	0.605 0.03%	0.398 0.05%	13.831 0.01%	0.296 0.06%
12	0.798 0.09%	5.416	1.536 0.03%	24.594 0.05%	0.617 0.06%	0.405 0.05%	14.039 0.00%	0.301 0.04%
13	0.810 0.03%	5.502	1.555 0.02%	25.106 0.04%	0.627 0.06%	0.410 0.07%	14.206 0.03%	0.305 0.01%
14	0.820 0.04%	5.571	1.570 0.03%	25.534 0.02%	0.636 0.04%	0.415 0.00%	14.345 0.01%	0.309 0.12%
15	0.829 0.06%	5.631	1.584 0.03%	25.867 0.01%	0.643 0.04%	0.419 0.01%	14.463 0.01%	0.312 0.09%

5.2. Performance Measures Used in Simulations

The first step is to compare Fundamental Scale with all other scales. Basically, a random PCM is generated using Fundamental Scale and the same PCM is also generated by using the corresponding numerical values in the measured scale. For instance, if Balanced Scale is measured, a score of "5" assigned by Fundamental Scale is converted to its correspondent "2.333" in Balanced Scale. A high number of PCM pairs are generated, and for each of them, several characteristics are measured. Then, a set of performance measures are calculated using all of the acceptably consistent PCMs, referred to as "trials". These performance measures are explained below:

- Measured Scale Lower CR: This performance measure represents the percentage of trials, which have lower CR values when generated by measured scale rather than Fundamental Scale.
- Fundamental Scale Lower CR: This performance measure represents the percentage of trials, which have lower CR values when generated by Fundamental Scale rather than measured scale.
- Equal CR: This performance measure represents the percentage of trials, which have the same CR values when generated by measured scale and Fundamental Scale.
- Fundamental Scale Inconsistent: This measure represents the percentage of trials, which are inconsistent when generated by Fundamental Scale but consistent when generated by measured scale.
- Measured Scale Inconsistent: This measure represents the percentage of trials, which are inconsistent when generated by measured scale but consistent when generated by Fundamental Scale.
- Both Scales Consistent: This measure represents the percentage of trials, which are consistent for both Fundamental Scale and measured scale.
- Same Best Chosen: This measure represents the percentage of trials, for which, the PCM generated by measured scale suggests the same element as the most important one.

- Same Worst Chosen: This measure represents the percentage of trials, for which, the PCM generated by measured scale suggests the same element as the least important one.
- Kendall’s Tau Mean: This measure represents the rank correlation of weight vectors for each PCM pair generated by Fundamental Scale and measured scale, in terms of Kendall’s Tau correlation coefficient for all trials.
- Tau Standard Deviation: This measure represents the standard deviation of Kendall’s Tau correlation coefficients for all pair of PCMs generated by Fundamental Scale and measured scale.
- Number of Deviations (Different Best): This measure represents the number of trials, for which, the PCM generated by measured scale and the PCM generated by Fundamental Scale suggest different elements as the most important one.

5.3. Simulation Steps and Results

The algorithm used to measure these metrics is basically as follows:

- Generate a random matrix using Fundamental Scale
- Generate the same matrix using the corresponding numerical values on the measured scale (see Table 23 for corresponding values in different scales)
- Check if CR of either matrix is below the CR limit given (step is repeated for CR_Limit=0.15, CR_Limit=0.10, and CR_Limit=0.05 for each scale)
- If either of the matrices (or both) is (are) below the given CR limit:
    - Calculate final weights of both matrices
    - Check similarity of final weight vectors using Kendall’s Tau
    - Check whether both matrices select the same best/worst alternatives, using Fundamental Scale as a benchmark
- Repeat the previous steps until the simulation finds 1000 consistent matrices (either one or both consistent):
    - For each measured scale
        - For each matrix size (from 3x3 to 7x7)

        - For each CR limit (0.15, 0.10, 0.05) (For 7x7 matrices, CR limit of 0.05 have not been simulated as it is estimated that simulations would take very long time.)
- For each CR limit and matrix size of the measured scale, calculate the percentage of the matrices:
    - Which have a lower CR than Fundamental Scale matrix
    - Which have a higher CR than Fundamental Scale matrix
    - Which have the same CR with Fundamental Scale matrix

At the end, for each row that will be presented in the following parts, we have obtained a pair (one formed by Fundamental Scale and one formed by the measured scale) of 1000 matrices, with at least one of the matrices is consistent for each pair.

Table 26 shows the results of different scales for 3x3 matrices at CR limit of 0.10. The first three rows sum up to 100% for each scale. They show the percentages of matrices based on their CR comparison. For Balanced Scale, for instance, 62.40% of the 1000 matrices would have lower CR values when these matrices are formed by using the Balanced Scale instead of Fundamental Scale. 37.60% of the matrices formed by using Fundamental Scale, on the other hand, would have lower CR values than those formed by using the Balanced Scale. The remaining, if existed, would be the ones that have the same CR with both scales.

Based on the first row, the results indicate that all measured scales, except for Logarithmic and Root Square scales, have better consistency characteristics compared to Fundamental Scale. While Power Scale seems superior to the others, Balanced, Exponential, Geometric and Inverse Linear scales seem somewhat similar in their consistency characteristics when compared with Fundamental Scale.

The 4^{th}, 5^{th} and 6^{th} rows also sum up to 100% and they represent the percentages of the cases where either of the scales are both generated consistent matrices based on the given CR limit. When Balanced Scale is compared with Fundamental Scale, it generated consistent pairwise comparison matrices in 34.10% of the cases while Fundamental Scale generated inconsistent pairwise comparison matrices. On the other hand, for the 21.20% of the cases, Fundamental Scale generated consistent pairwise comparison matrices while Balanced Scale generated inconsistent pairwise comparison matrices. In the remaining 44.70% of the cases, both scales generated consistent pairwise comparison matrices.

Table 26 Performance Measurement of Different Scales

Performance Measures (3x3 Matrices, CR Limit=0.10)	Balanced	Exponential	Geometric	Inverse Linear	Logarithmic	Power	Root Square
Measured Scale Lower CR	62.40%	68.90%	72.60%	63.50%	34.50%	98.30%	1.00%
Fundamental Scale Lower CR	37.60%	31.10%	27.40%	36.50%	65.50%	1.70%	98.70%
Equal CR	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.30%
Fundamental Scale Inconsistent	34.10%	41.00%	36.00%	41.60%	6.50%	27.30%	2.90%
Measured Scale Inconsistent	21.20%	11.90%	19.50%	22.20%	20.90%	0.40%	8.60%
Both Scales Consistent	44.70%	47.10%	44.50%	36.20%	72.60%	72.30%	88.50%
Same Best Chosen	96.70%	97.40%	99.20%	95.90%	99.60%	99.60%	99.50%
Same Worst Chosen	97.20%	97.00%	98.50%	96.30%	100.00%	99.90%	99.30%
Kendall's Tau Mean	0.9576	0.9608	0.9821	0.9457	0.9953	0.9963	0.9902
Tau Standard Deviation	0.1653	0.1540	0.1019	0.1857	0.0461	0.0477	0.0747
Number of Deviations (Different Best)	33	26	8	41	4	4	5

The 7th and 8th rows indicate the percentage of matrices, for which, both Fundamental Scale and measured scale selected the same comparison elements as the best and worst alternatives, respectively. The results show that all scales suggested similar best and worst alternatives, though Inverse Linear Scale showed poor performance in this measure, when compared to the other scales.

In parallel to the same best and worst measures, weight vectors of all scales were evaluated and compared with those of Fundamental Scale. Vector comparison is carried out by Kendall's Tau (Kendall, 1938) approach. The 9th and 10th rows show that all scales seem to have suggested mostly the same rankings for compared elements as Fundamental Scale does. Inverse Linear Scale, however, is below the average with its $\tau = 0.9457$ and the largest standard deviation of 0.1857.

Lastly, the 11th row shows the number of matrices (out of 1000 matrices), for which, the best alternative suggested by the measured scale is different than the best alternative suggested by Fundamental Scale. Inverse Linear and Balanced scales seem to be the ones that deviate the most from Fundamental Scale in this respect. While Exponential Scale shows moderate deviations, Geometric, Logarithmic, Power, and Root Square scales seem to be mostly parallel to Fundamental Scale.

Individual performances of each scale in each matrix dimension and CR limit can be seen in Appendix B. Figure 12 illustrates the scale performances for "CR limit=0.1" (Saaty's CR limit), based on the data given in Appendix B. Apparently, Power Scale dominates all other scales with respect to the percentage of generated matrices which have lower CR compared to Fundamental Scale. Geometric and Exponential scales follow Power Scale and converge to its performance at large scale matrices. While Inverse Linear and Balanced scales also seem to have an increasing performance with the matrix size, Logarithmic and Root Square scales have very poor performance, especially in larger matrices.

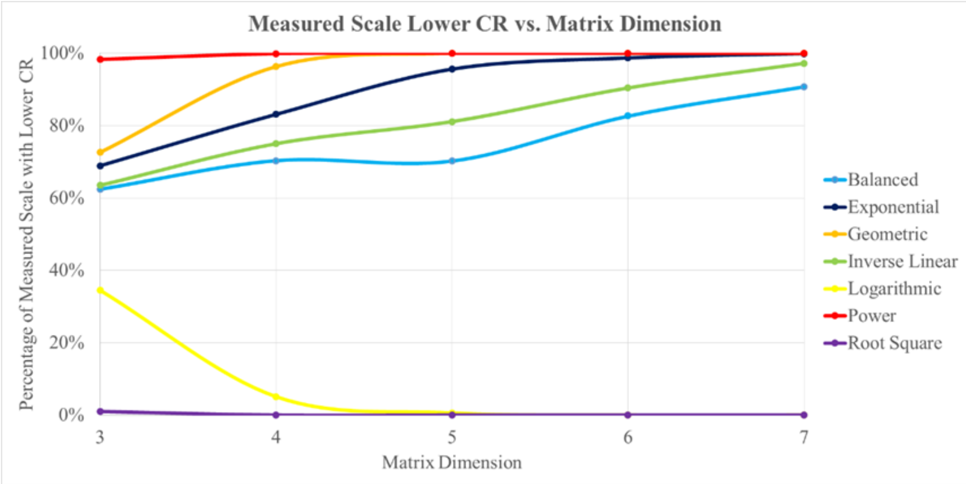

Figure 12 Comparison of scales based on the CR of generated matrices

Figures 13 and 14, show the performances of scales based on "same best" and "same worst", when compared with Fundamental Scale. Logarithmic, Power and Root Square scales appear to have the best performance in terms of "same best" and "same worst" while Balanced, Exponential and Geometric scales have lower performances. Inverse Linear Scale, on the other hand, seems to have the worst performance although its worst performances in "same best" and "same worst" are approximately 82% and 84%, respectively.

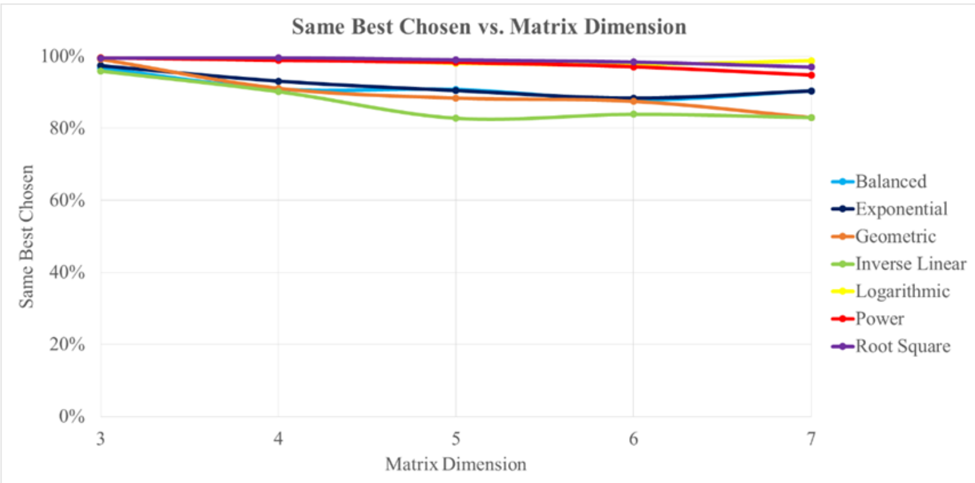

Figure 13 Comparison of scales based on same best selection

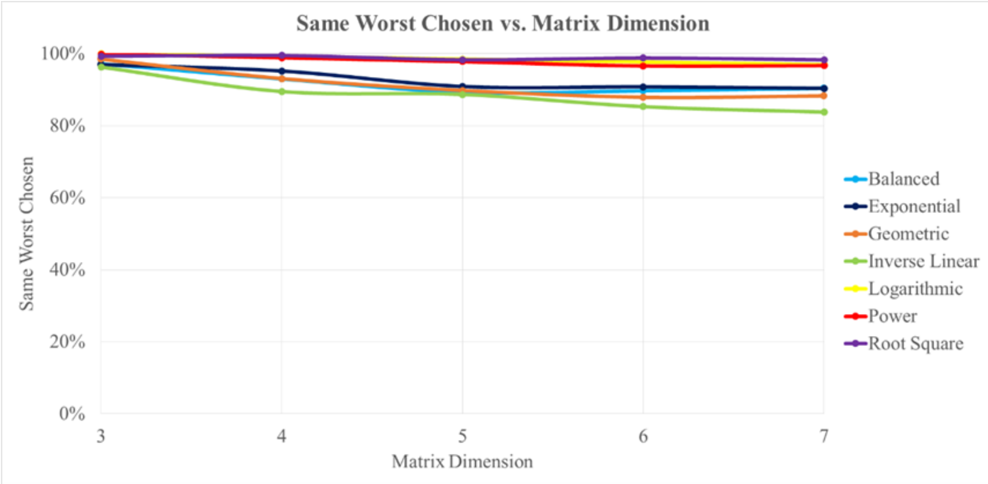

Figure 14 Comparison of scales based on same worst selection

Another important performance measure is the Kendall's Tau values of the generated weight vectors compared to those generated by using Fundamental Scale. Figure 15 shows that Kendall's Tau performances are somewhat parallel to the "same best" and "same worst" performances, given in Figures 13 and 14. That is Logarithmic, Power and Root Square scales have the best performance in terms of Kendall's Tau criteria. While Balanced, Exponential and Geometric scales have lower performance, Inverse Linear Scale again has the worst performance in a weight-vector-generation-related criterion.

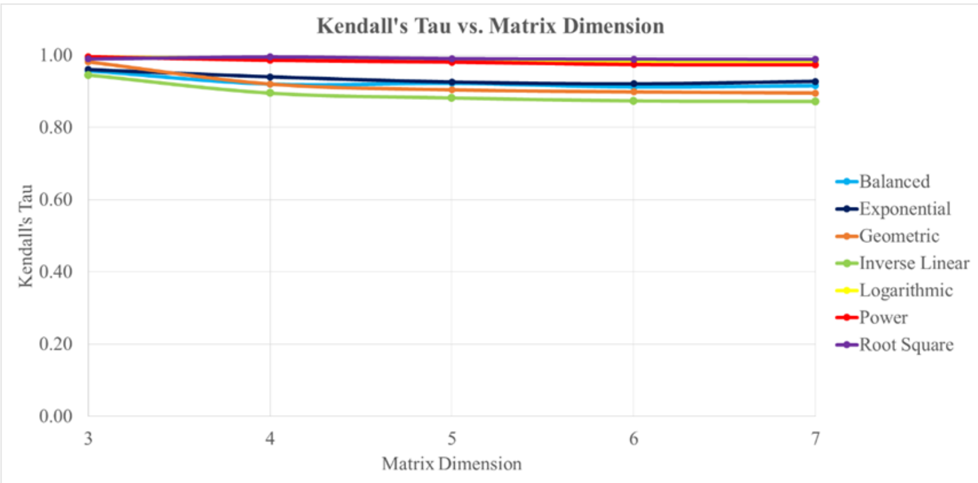

Figure 15 Comparison of scales based on Kendall's Tau performance

As all scales have different strengths and weaknesses in each comparison criterion, an overall comparison of them may be beneficial to clarify the advantages and disadvantages of scales in different criterion, and even draw more general conclusions based on the nature of the numbers in scales by clustering them in groups.

Table 27 shows the average performances of scales in terms of performance masures (the average of all matrix sizes at CR limit=0.10). Some of the performance measures reported Table 26 were aggregated to indicate more accurate results. “Measured Scale Lower CR” is taken alone as it can show the performance of the measured scale in terms of CR. Additionally, the cases where the measured scale generated consistent results (CR<0.10) are aggregated by extracting the inconsistent cases (Measured Scale Inconsistent) from 100%. The remaining measures (Same Best Chosen, Same Worst Chosen, Kendall’s Tau Mean, Tau Standard Deviation, and Number of Deviations (Different Best) are directly averaged as they show significant performance indications.

Table 27 Average performances of scales in terms of performance measures (all matrix sizes, CR limit=0.10)

SCALE	Measured Scale Lower CR	Measured Scale Consistent	Same Best Chosen	Same Worst Chosen	Kendall's Tau Mean	Tau Standard Deviation	Number of Deviations (Different Best)	Percent Deviation for Different Best
Balanced	73.48%	79.93%	94.38%	94.95%	95.61%	90.97%	94.38%	96.40%
Exponential	85.43%	92.75%	94.60%	95.13%	95.62%	91.31%	94.60%	96.38%
Geometric	90.85%	93.70%	93.13%	93.70%	94.51%	90.35%	93.13%	96.25%
Inverse Linear	78.40%	83.13%	91.03%	91.73%	92.76%	86.77%	91.03%	94.49%
Logarithmic	14.35%	45.60%	99.00%	99.08%	99.19%	95.41%	99.00%	99.56%
Power	99.20%	99.78%	98.60%	98.53%	98.90%	94.55%	98.60%	99.44%
Root Square	0.28%	61.90%	99.35%	99.35%	99.27%	95.62%	99.35%	99.87%

Some of the measures (Measured Scale Lower CR, Masured Scale Consistent, Same Best Chosen, Same Worst Chosen, Kendall's Tau Mean) can be used directly to show the scales' performances, as the higher they are, the better the scale. However, for other measures (Tau Standard Deviation and Number of Deviations -Different Best, the lower the value in Table 27, the better the scale.

The map chart in Figure 16 illustrates the performance scores of different scales based on the directly usable data shown in Table 27. For all scales, the closer the point of the scale to the outer edge, the better the scale in terms of the corresponding performance measures. Apparently, Power Scale is very close to the outer edge in all of the measures, thus, seems to be the best scale in terms of our performance measures. Geometric, Exponential, Inverse Linear, and Balanced Scales follow the Power Scale, respectively. Although Logarithmic and Root Square scales show comparable performances in other performance measures, their performances in "Measured Scale Lower CR" and "Measured Scale Consistent" are considerably lower compared to those of other scales. Therefore, in terms of these performance measures, Logarithmic and Root Square scales are very weak when compared with Fundamental Scale. An important point to mention here is that the scales can be divided to three main groups, based on the performances in consistency-related measures (Measured Scale Lower CR and Measured Scale Consistent). The first group consists of Exponential, Geometric, and Power scales that come to the forefront with their scores. A common point of these three scales is that all of them have an upper limit greater than 9.000, which was originally proposed by Saaty (1980). The second group consists of Balanced and Inverse Linear scales with the upper limit of 9.000 for both scales. The third group consists of Logarithmic and Root Square scales with the upper limits of 3.322 and 3.000, respectively. From this point of view, it appears that the upper-bound-related inconsistency issue is supported by the upper limit of the scale, and becomes more critical to the scales which have upper limits less than 9.000. Conversely, for those scales with upper limits greater than 9.000, it can be said that the inconsistency issue is far less.

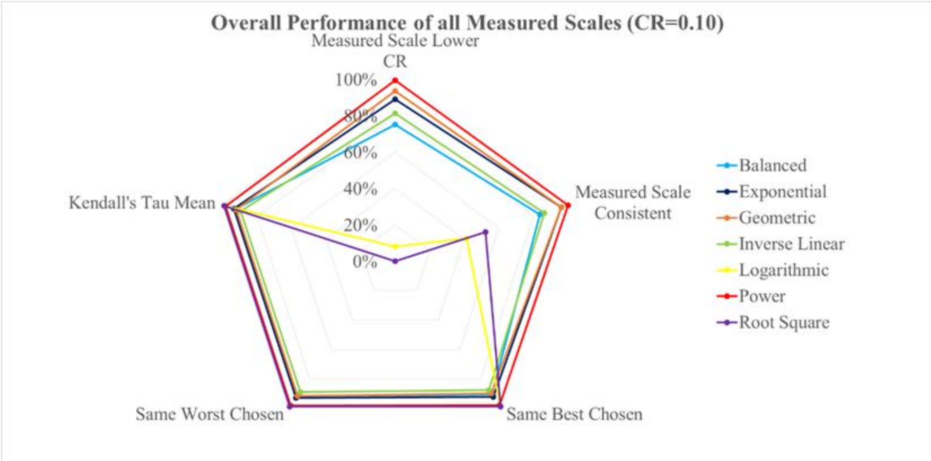

Figure 16 Comparison of scales based on estimated performance scores (CR=0.10)

So far, we were only concerned with the performances and average scores when CR limit is taken as "0.10". Our simulations, however, included the cases where CR limit is changed to "0.05" and "0.15" in order to analyze the sensitivity of results based on CR limit. Figure 17 and 18 show that there is no significant change in the performance of scales when CR limit is changed. Therefore, we may contently say that the previously stated comments hold even when CR limit is different than "0.10".

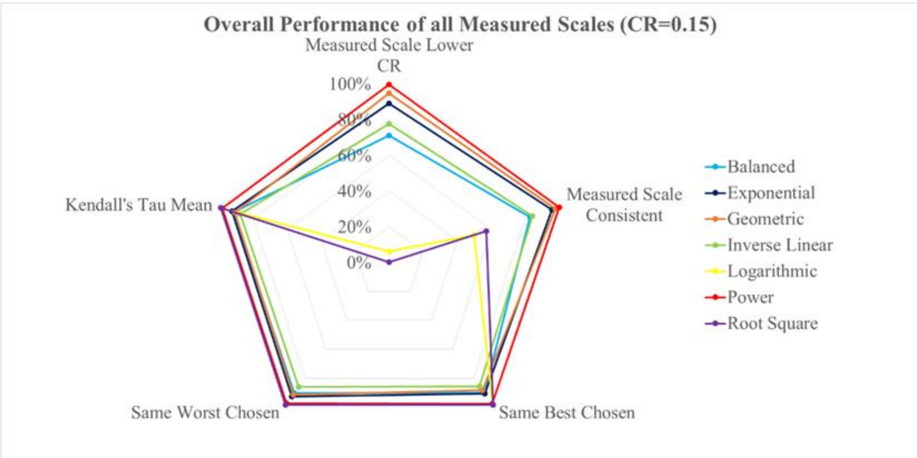

Figure 17 Comparison of scales based on estimated performance scores (CR=0.15)

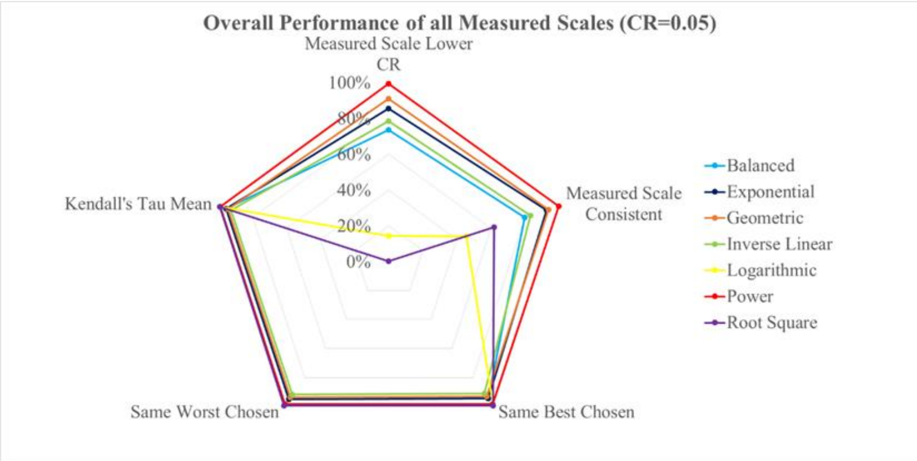

Figure 18 Comparison of scales based on estimated performance scores (CR=0.05)

CHAPTER 6

COMPARISON OF PAIRWISE COMPARISON SCALES ON A SAMPLE DECISION PROBLEM

Until now, we have only discussed the theoretical side of the pairwise comparison scales, including the Exponential Scale proposed by this study. At this point, an illustrative example would be beneficial to show where Exponential Scale is positioned among the existing pairwise comparison scales. Franek and Kresta (2014) studied this topic on an example based on Saaty's study (2003), where an example involving the prioritization of criteria used to buy a house for a family. The initial PCM provided by the family members, which is known to be currently inconsistent, is shown in Table 28.

Table 28 Inconsistent PCM (S) for the decision of buying a house for a family

Matrix S	Size	Transportation	Neighborhood	Age	Yard	Modern	Condition	Finance
Size	1	5	3	7	6	6	1/3	1/4
Transportation	1/5	1	1/3	5	3	3	1/5	1/7
Neighborhood	1/3	3	1	6	3	4	6	1/5
Age	1/7	1/5	1/6	1	1/3	1/4	1/7	1/8
Yard	1/6	1/3	1/6	3	1	1/2	1/5	1/6
Modern	1/6	1/3	1/4	4	2	1	1/5	1/6
Condition	4	5	1/6	7	5	5	1	1/2
Finance	4	7	5	8	6	6	2	1

Source: Saaty (2003)

As the first PCM, S, is known to be inconsistent, Saaty (2003) suggests a solution to improve the consistency by locating the most inconsistent pairwise comparison judgement and ask the DM to revise this judgement. This step can be carried out for other highly inconsistent pairwise comparison judgements, until the DM provides a PCM that has an acceptable CR (CR ≤ 0.1 in our example). For this example, Saaty locates the most inconsistent pairwise comparison judgement and changes it to a scale-defined score that is closest to this judgement's most consistent state. He calculates that a_{37}=6 shoud be as close to "1/2.18" as possible. Therefore, he hypothetically changes a_{37} to "1/2". From the reciprocity property, a_{73} is automatically changed to "2". This way, Saaty forms the consistent PCM for this problem, denoted as S'. The consistent matrix S' is shown in Table 29.

Table 29 Consistent PCM (S') for the decision of buying a house for a family

Matrix S'	Size	Transportation	Neighborhood	Age	Yard	Modern	Condition	Finance
Size	1	5	3	7	6	6	1/3	1/4
Transportation	1/5	1	1/3	5	3	3	1/5	1/7
Neighborhood	1/3	3	1	6	3	4	1/2	1/5
Age	1/7	1/5	1/6	1	1/3	1/4	1/7	1/8
Yard	1/6	1/3	1/6	3	1	1/2	1/5	1/6
Modern	1/6	1/3	1/4	4	2	1	1/5	1/6
Condition	4	5	2	7	5	5	1	1/2
Finance	4	7	5	8	6	6	2	1

Franek and Kresta (2014) note that although Saaty (2003) used the EVM in his study, they utilized the most frequent type of Row Geometric Mean Method (RGMM, previously described as LLSM), formulized as follows:

$$w_i = \frac{\left(\prod_{j=1}^{i} a_{ij}\right)^{1/i}}{\sum_{i=1}^{j} \left(\prod_{j=1}^{i} a_{ij}\right)^{1/i}}$$

CI and CR values are also calculated based on the previously introduced formulae, where RI values are used as the values previously given in Table 25.

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

Franek and Kresta (2014) used Microsoft Excel® to estimate weights, CI, and CR values. In this study, however, MATLAB® software is used to generate the scales and respective matrices to calculate weights, CI, and CR values. Therefore, obtained results are different than those of Franek and Kresta (2014). Table 30 and Table 31 show the results obtained for all scales, and tabulated similar to how Franek and Kresta (2014) did in their study.

Table 30 Weights and consistency measures for different pairwise comparison scales on inconsistent matrix S

Criteria Used in Matrix S	Balanced	Exponential	Fundamental	Geometric	Inverse Linear	Logarithmic	Power	Root Square
Size	0.174	0.186	0.175	0.168	0.166	0.164	0.146	0.162
Transportation	0.086	0.041	0.063	0.019	0.096	0.092	0.019	0.097
Neighborhood	0.146	0.130	0.149	0.100	0.142	0.152	0.105	0.150
Age	0.042	0.009	0.019	0.002	0.056	0.046	0.002	0.054
Yard	0.063	0.020	0.033	0.007	0.078	0.062	0.005	0.070
Modern	0.072	0.027	0.042	0.010	0.086	0.073	0.009	0.080
Condition	0.158	0.155	0.168	0.129	0.151	0.162	0.133	0.158
Finance	0.259	0.432	0.351	0.564	0.224	0.248	0.582	0.229
$\boldsymbol{\lambda_{max}}$	8.250	10.037	9.505	14.388	8.107	8.507	17.826	8.316
CI	0.036	0.291	0.215	0.913	0.015	0.072	1.404	0.045
CR	0.050	0.061	0.175	0.044	0.028	0.197	0.110	0.165

Table 31 Weights and consistency measures for different pairwise comparison scales on consistent matrix S'

Criteria Used in Matrix S'	Balanced	Exponential	Fundamental	Geometric	Inverse Linear	Logarithmic	Power	Root Square
Size	0.173	0.181	0.172	0.160	0.165	0.163	0.137	0.161
Transportation	0.086	0.040	0.062	0.018	0.096	0.092	0.018	0.097
Neighborhood	0.123	0.088	0.107	0.057	0.126	0.125	0.053	0.127
Age	0.042	0.009	0.019	0.002	0.056	0.046	0.002	0.054
Yard	0.062	0.019	0.032	0.006	0.078	0.062	0.005	0.070
Modern	0.072	0.026	0.042	0.010	0.086	0.072	0.008	0.079
Condition	0.185	0.217	0.224	0.208	0.170	0.194	0.232	0.184
Finance	0.257	0.420	0.343	0.539	0.224	0.246	0.546	0.228
$\boldsymbol{\lambda_{max}}$	8.017	8.407	8.745	8.976	7.988	8.296	11.475	8.160
CI	0.002	0.058	0.106	0.139	-0.002	0.042	0.496	0.023
CR	0.003	0.012	0.076	0.007	-0.003	0.115	0.039	0.084

The difference between the results of Franek and Kresta (2014) can be mainly attributed to the difference between RI values used in calculations. The results shown in Table 30 and Table 31 are calculated using our RI values previously given in Table 25.

CI and CR values are calculated both by using MATLAB® and Excel®. Excel® estimation of CI values is carried out by using methodology defined by Mu and Pereyra-Rojas (2017):

- Calculate weights of criteria (by LLSM in our case)
- Add transpose of the weight vector as the top row of the original PCM

- Multiply each element on the first column of the original PCM with the first element of the transpose weight vector. Then repeat this step for all remaining columns.
- Sum up all the rows and write the results to the right of the multiplied PCM, so that results form a column vector with the size of (8x1).
- Put the (8x1) column vector to the left of the original weight vector. Then divide the first-row element of the (8x1) column vector by the first-row element of the original weight vector. Repeat the same step for all remaining rows to obtain a (8x1) division vector.
- Take the average of the elements of (8x1) division vector as an estimate of CI.

As MATLAB® and Excel® results for CI values are very close to each other, we used Excel® results so that the reader can easily check the values without any sophisticated software.

As Franek and Kresta (2014) mentioned, Geometric and Power scales yield a higher score for the most important criterion. Similarly, as expected, Exponential Scale shows the same characteristic to emphasize the value of the most important criterion. In the former case (inconsistent matrix S), Balanced, Exponential, Geometric, and Inverse Linear scales seem to be consistent while the Fundamental, Logarithmic, Power, and Root Square scales are inconsistent. In the latter case (consistent matrix S'), however, all scales except for Inverse Linear and Logarithmic scales seem to be consistent. Surprisingly, in both MATLAB® and Excel® calculations, the $\boldsymbol{\lambda_{max}}$ value for the matrix generated using the Inverse Linear Scale was calculated as “7.988”, which is lower than the matrix dimension “8”. Therefore, considering the CI definition, the numerator of the fraction becomes negative and it results in a negative CR value. This brings the question to the mind whether CI and CR are proper means of checking the inconsistency of a PCM.

In this particular case, it is seen that Exponential and Geometric scales are consistent in both cases, where the Fundamental and Power scales are consistent in only one.

Table 32 Percent changes between S and S' cases

Percent (%) Changes Between Cases	Balanced	Exponential	Fundamental	Geometric	Inverse Linear	Logarithmic	Power	Root Square
Size	**0.60**	**2.75**	**2.08**	**4.56**	**0.30**	**0.73**	**6.16**	**0.51**
Transportation	**0.60**	**2.75**	**2.08**	**4.56**	**0.30**	**0.73**	**6.16**	**0.51**
Neighborhood	**15.50**	**32.24**	**28.22**	**43.25**	**11.23**	**17.63**	**49.58**	**14.82**
Age	**0.60**	**2.75**	**2.08**	**4.56**	**0.30**	**0.73**	**6.16**	**0.51**
Yard	**0.60**	**2.75**	**2.08**	**4.56**	**0.30**	**0.73**	**6.16**	**0.51**
Modern	**0.60**	**2.75**	**2.08**	**4.56**	**0.30**	**0.73**	**6.16**	**0.51**
Condition	16.93	39.58	33.59	60.51	11.97	19.64	74.65	16.21
Finance	**0.60**	**2.75**	**2.08**	**4.56**	**0.30**	**0.73**	**6.16**	**0.51**
$\boldsymbol{\lambda_{max}}$	**2.83**	**16.24**	**7.99**	**37.61**	**1.48**	**2.48**	**35.63**	**1.87**
CI	**93.10**	**80.00**	**50.47**	**84.72**	**111.58**	**41.59**	**64.64**	**49.22**
CR	**93.10**	**80.00**	**50.47**	**84.72**	**111.58**	**41.59**	**64.64**	**49.22**

Table 32 shows the percent changes in the weights, $\boldsymbol{\lambda_{max}}$, CI and CR values for each scale. Bold black values show there was a decrease when S is changed to S', and others show that there was an increase. All percent changes are calculated by the following rule:

- The former case numerical value is subtracted from the latter case numerical value
- Result is divided by the former case numerical value

From this analysis, it can be concluded that the greatest change in CR was observed for Inverse Linear Scale. The negative value calculated previously also has a supplementary effect to this significant change. Balanced, Exponential and Geometric scales also showed a significant increase in CR values. The Fundamental, Logarithmic, Power, and Root Square scales, on the other hand, remained below the others in terms of percent CR change.

Apparently, percent deviations tend to increase when Power and Geometric scales are used. These deviations are moderate in Exponential and Fundamental Scale cases, unlike comparatively low deviations in Balanced, Inverse Linear, Logarithmic, and Root Square scale cases. Although these deviations can be observed from all criteria, they are particularly more significant in "Condition" and "Neighborhood" as these criteria are evaluated with changed numerical scores.

We believe consistency sensitivity is not directly related to the CR value, but the percent deviation between the cases. Variance of weights is already based on change. Based on these results, the conclusion of Franek and Kresta (2014) may be revised as shown in Table 33.

Table 33 Classification of judgement scales based on consistency and allocation of weights

	Consistency Sensitivity	**Variance of Weights**
High	Inverse Linear	Geometric Power
Moderate	Balanced Exponential Geometric	Exponential Fundamental
Low	Fundamental Logarithmic Power Root Square	Balanced Logarithmic Power Root Square

CHAPTER 7

CONCLUSIONS AND FINAL REMARKS

Real life decision problems rarely depend on a single parameter. On the contrary, they are often complex problems, which involve multiple alternatives, criteria, and stakeholders. AHP is a very common and powerful multi-criteria decision making tool used in numerous business sectors by managerial decision makers (DM). The method is applied by verbally comparing alternatives and converting these verbal comparisons to numerical scores. The power of AHP is based on three main sources:

- Its systematic process of dividing complex decision problems into smaller manageable parts,
- Using pairwise comparison of elements rather than considering all at once, and
- Providing means to systematically check DMs consistency in pairwise comparisons.

Consistency concept is critical for managerial DMs in complex decision problems, as better consistency means more structured and well-made pairwise comparisons. On the other hand, if DM is not adequately consistent, then pairwise comparisons should be revised in order to make sure all of them reflect the correct evaluations of the DM. This revision becomes more time consuming and costly for managerial DMs as the complexity level of the decision problems increases.

Consistency in AHP is mainly attributed to the transitivity axiom. Transitivity can be divided into two:

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

According to the transitivity axiom of AHP, a set of pairwise comparisons are fully consistent if and only if all of them are cardinally transitive. However, the scales suggested for and used in AHP tend to be cardinally intransitive.

If we approach inconsistency issue from the perspective of numerical pairwise comparison scales, two kinds of inconsistency can be defined. The first one is due to the fact that a numerical pairwise comparison scale has upper and lower limits, which make some combination (multiplication) of numerical scores cardinally intransitive. In this study, it is shown that most of the existing scales are significantly affected by this type of inconsistency characteristic. However, it is noted that any type of scale with an upper and lower bound would suffer from such inconsistency. More importantly, these scales (except for Geometric Scale) are significantly affected by the second type of inconsistency, which is due to the discreteness of the scale.

In order to reduce inconsistencies based on cardinal intransitivity and scale discreteness, a new scale based on Fibonacci Series is proposed in this study, which is called “Exponential Scale”. Unlike most of the other scales, Exponential Scale is a continuous function, considering that reciprocals of all members of the scale are still a member of the same scale. As expected, Exponential Scale appeared more advantageous when the two types of scale-based inconsistency are concerned.

By detailed simulations, previously proposed scales (except for Fundamental Scale) and Exponential Scale have been compared in this study with Fundamental Scale using some performance measures. In these simulations, especially when consistency-related performance measures are taken into account, Power Scale appeared better than the other scales. While Exponential and Geometric scales followed Power Scale closely, Inverse Linear and Balanced scales showed only moderate performance. Logarithmic and Root Square scales, on the other hand, showed very poor performance with respect to consistency-related performance measures, when compared to the most widely used Fundamental Scale. It is notable that scales with a limit greater than 9 (Exponential, Geometric, and Power scales) have better performances compared to the scales with an upper limit of 9 (Inverse Linear and Balanced scales). Scales having upper limits less than 9 (Logarithmic and Root Square scales) have the lowest performances. Therefore, we can conclude that the upper limit of a numerical pairwise comparison scale is actually a significant parameter of consistency, and in general, scales with greater limits than 9 appear to be more consistent in numerical analysis of the same verbal judgement sets. Moreover, this is an indication of the suggestion that AHP favors wider scales in terms of consistency.

At the last step, all scales are applied on the same sample decision problem in inconsistent and consistent cases (based on Fundamental Scale), and their characteristics were evaluated. Although, we believe, a single example is not adequate to draw general conclusions, the scales with higher upper bounds (Exponential, Geometric, and Power) yielded a higher weight for the most important criterion, as expected. Furthermore, Exponential and Geometric scales generated consistent PCMs for both cases while the Fundamental and Power scales generated a consistent matrix in only one of the cases.

Although Geometric Scale has a similar performance to the proposed Exponential Scale, we agree with the idea of Ji and Jiang (2003) that the use of a scale with an upper limit of 256 severely violates the homogeneity axiom of AHP. Therefore, we believe that the use of Exponential Scale would be much more appropriate than Geometric Scale. Power Scale violates the same axiom more than Exponential Scale. Therefore, in this regard, Exponential Scale appears to be superior to the other scales.

Another important finding is related to the performances of Exponential Scale and Power Scale's for larger PCMs. Although this study did not include sizes larger than 7x7 PCMs, the trends of performances indicate that both scales would have very similar performances for larger PCMs. This finding further supports Exponential Scale, especially for larger scale matrices, considering that Power Scale violates the homogeneity axiom more than Exponential Scale.

Power Scale was proposed by Harker and Vargas (1987), and was criticized in the same study by using a single example application of AHP. Harker and Vargas (1987) used the same verbal comparison scale with different numerical pairwise comparison scales and generated the final weights to compare them with the actual normalized distances between Philadelphia and the other cities. They concluded that the weight vector generated by using Fundamental Scale yielded the highest correlation with the actual normalized distances. However, we believe that they missed an important point. A rational DM, who knows what numerical value corresponds to his/her verbal judgements, may not use the same verbal evaluations for the Fundamental and Power scales. Indeed, provided the numerical scale is known, a DM may adjust himself/herself so that the numerical judgements represent his/her actual idea of the ratio weights (distances in the sample problem). If DMs are asked questions about measurable elements, about which they at least have some preliminary information, they may assess numerical ratios more accurately. Thus, we believe that the assumption of the same verbal scale used should be approached carefully in AHP studies. In this regard, we believe that Power Scale may not be as disadvantageous as Harker and Vargas (1987) claimed to be, and it may bring the advantage of more consistent PCMs which decreases the burden of revising inconsistencies.

The same limitation applies to our study as well. In our simulations, we generated PCMs first based on Fundamental Scale, then directly converted the numbers to respective numbers in other scales. That is, we used the same verbal scale for both the Fundamental and the measured scale. A further study may be to assess real DMs' perception of the verbal scale when different numerical scales are used. That is, the following question should be investigated: Does a DM change the verbal judgement if the numerical scale is changed?

Another limitation of our study is that we only used EVM to generate weight vectors. In our preliminary simulations, we confirmed that the Approximate Method (a.k.a. the Rule of Thumb) yielded very close results to those of EVM, and as result, we decided to continue only with EVM for the sake of computational limitations. However, LLSM (RGMM) has been attracting more and more attention for the past two decades, and the simulations could have also been run for LLSM as the weight derivation method.

Another limitation of our study is that we only used EVM to assess the consistencies of matrices. It has been mentioned that GCI can also be an alternative way for consistency measurement. Further studies may evaluate the combinations of weight derivation methods (EVM and LLSM) with consistency measurement methods (EVM and GCI).

Although we simulated a general comparison of numerical pairwise comparison scales, we are aware that we only represented the results of the PCMs, which have lower CR values than the CR limit than we designated. A complete mapping of numerical pairwise comparison scale behavior throughout all CR range may broaden the knowledge of the literature, and thus, may be considered as a possible future study topic.

We believe our study is different than the others, since it does not simply compare numerical pairwise comparison scales on only single decision problem, but it analyzes the simulation of many cases and compares the other scales with the most widely used Fundamental Scale. We also believe that the inconsistency characteristics expressed by Ji and Jiang (2003) should have been elaborated on, as their preliminary analysis was not very detailed. In our study, we believe that we showed these characteristics better with our detailed analysis of every single possible numerical pairs in all scales.

Moreover, our study brings a different approach to consistency sensitivity concept, which was suggested by Franek and Kresta (2014). We believe that consistency sensitivity should be measured by the deviation between the consistent and inconsistent cases of the given sample problem. That is, consistency sensitivity should be regarded as the percent change in the CR when an element (or a set of elements) are changed in a PCM.

Finally, we would like to emphasize that general conclusions should not be drawn based only on a single application of AHP. As our simulations showed, although they do not represent the majority, there are significantly many cases where Fundamental Scale yields more consistent results than Exponential, Geometric or Power scales. In fact, a single example may well be one of these cases, where Fundamental Scale seems superior to all others. Therefore, we believe that simulations involving many possible cases would provide more reliable results.

REFERENCES

Abdi, M. R. & Labib, A., 2003. A Design Strategy for Reconfigurable Manufacturing Systems (RMSs) Using Analytical Hierarchical Process (AHP): A Case Study. *International Journal of Production Research,* 41(10), pp. 2273-2299.

Aguaron, J. & Moreno-Jiménez, J. M., 2003. The Geometric Consistency Index: Approximated Thresholds. *European Journal of Operational Research,* Volume 147, pp. 137-145.

Al-Harbi, K. M.-S., 2001. Application of the AHP in Project Management. *International Journal of Project Management,* 19(1), pp. 19-27.

Alonso, J. A. & Lamata, M. T., 2004. Estimation of the Random Index in the Analytic Hierarchy Process. *Proceedings of Information Processing and Management of Uncertainty in Knowledge- Based Systems,* pp. 317-322.

Alonso, J. A. & Lamata, M. T., 2006. Consisteny in the Analytic Hierarchy Process: A New Approach. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* 14(4), pp. 445-459.

Amiri, M., Zandieh, M., Soltani, R. & Vahdani, B., 2009. A Hybrid Multi-Criteria Decision-Making Model for Firms Competence Evaluation. *Expert Systems with Applications,* Volume 36, pp. 12314-12322.

Badri, M., 2001. Combining the AHP and GP Model for Quality Control Systems. *International Journal of Production Economics,* Volume 72, pp. 27-40.

Bana e Costa, C. A., De Corte, J. M. & Vansnick, J. C., 2012. MACBETH. *International Journal of Information Technology and Decision Making,* 11(2), pp. 359-387.

Bana e Costa, C. A. & Vansnick, . J.-C., 2008. A Critical Analysis of the Eigenvalue Method Used to Derive Priorities in AHP. *European Journal of Operational Research,* Volume 187, p. 1422–1428.

Barzilai, J. & Golany, B., 1997. Deriving Weights from Pairwise Comparison Matrices. *Journal of the Operational Research Society,* Volume 48, pp. 1226-1232.

Bhushan, N. & Rai, K., 2004. Strategic Decision Making. In: *Applying the Analytic Hierarchy Process.* London: Springer-Verlag London, pp. 11-21.

Blaira, A. R., Nachtmann, R., Saaty, T. L. & Whitaker, R., 2002. Forecasting the Resurgence of the US Economy in 2001: An Expert Judgement Approach. *Socio-Economic Planning Sciences,* 36(2), pp. 77-91.

Brans, J.-P., 1982. *L'ingénièrie de la décision; Elaboration d'instruments d'aide à la décision. La méthode PROMETHEE.* Québec, Presses de l'Université Laval, pp. 183--213.

Budescu, D. V., Zwick, R. & Rapoport, A., 1986. A Comparison of the Eigenvalue Method and the Geometric Mean Procedure for Ratio Scaling. *Applied Psychological Measurement,* Volume 10, pp. 69-78.

Cheng, C.-H., 1997. Evaluating Naval Tactical Missile System by Fuzzy AHP Based on Grade Value of Membership Function. *European Journal of Operational Research,* 96(2), pp. 343-350.

Chen, S.-J. (. & Lin, L., 2003. Decomposition of Interdependent Task Group for Concurrent Engineering. *Computers & Industrial Engineering,* 44(3), pp. 435-459.

Choo, E. & Wedley, W., 2008. Comparing Fundamentals of Additive and Multiplicative Aggregation in Ratio Scale Multi-Criteria Decision Making. *The Open Operational Research Journal,* Volume 2, pp. 1-7.

Crawford, G. B., 1987. The Geometric Mean Procedure for Estimating the Scale of a Judgement Matrix. *Mathematical Modeling,* Volume 9, pp. 327-334.

Crawford, G. & Williams, C., 1985. A Note on the Analysis of Subjective Judgement Matrices. *Journal of Mathematical Psychology,* Volume 29, pp. 387-405.

Davoodi, A., 2009. On Inconsistency of a Pairwise Comparison Matrix. *International Journal of Industrial Mathematics,* 1(4), pp. 343-350.

Dijkstra, T. K., 2013. On the Extraction of Weights from Pairwise Comparison Matrices. *Central European Journal of Operations Research,* 21(1), pp. 103-123.

Dodd, F. J. & Donegan, H. A., 1994. Comparison of Prioritization Techniques Using Interhierarchy Mappings. *Journal of the Operational Research Society,* 46(4), pp. 492-498.

Dodd, F. J., Donegan, H. A. & McMaster, T. B. M., 1992. A New Approach to AHP Decision-Making. *Journal of the Royal Statistical Society. Series D (The Statistician),* 41(3), pp. 295-302.

Dong, Y., Xu, Y., Li, H. & Dai, M., 2008. A Comparative Study of the Numerical Scales and the Prioritization Methods in AHP. *European Journal of Operational Research,* 186(1), pp. 229-242.

Ehie, I. C. & Benjamin, C. O., 1993. An Integrated Multi-Objective Planning Model: A Case Study of Zambian Copper Mining Industry. *European Journal of Operational Research,* 68(2), pp. 160-172.

El Hefnawy, A. & Mohammed, A. S., 2014. Review Of Different Methods For Deriving Weights In The Analytic Hierarchy Process. *International Journal of the Analytic Hierarchy Process,* 6(1).

Fashoto, S. G., Owolabi, O. & Adekoya, A., 2016. Integration Of Geometric Mean Induced Bias Matrix Into Analytical Hierarchy Process To Enhance Decision Making Model For Supplier Selection In Healthcare Service Delivery. *International Journal of Business and Management Studies,* 5(2), pp. 429-436.

Fechner, G. T., 1860. *Elemente der Psychophysik.* Leipzig: Breitkopf und Hârtel.

Forman, E. H., 1990. Random Indices for Incomplete Pairwise Comparison Matrices. *European Journal of Operational Research,* 153-155.48(1).

Franek, J. & Kresta, A., 2014. Judgment Scales and Consistency Measure in AHP. *Procedia Economics and Finance*, 12(1), pp. 164-173.

French, S., 1988. *Decision Theory: An Introduction to the Mathematics of Rationality.* West Sussex: Ellis Horwood Ltd.

Gigerenzer, Gerd; Todd, P. M.; The ABC Research Group, 1999. Simple Heuristics That Make Us Smart. In: *Evolution and Cognition.* New York: Oxford University Press, pp. 3-34.

Goepel, K. D., 2018. Comparison of Judgment Scales of the Analytical Hierarchy Process - A New Approach. *International Journal of Information Technology and Decision Making,* 20 05.

Golden, B. L. & Wang, Q., 1990. An Alternative Measure of Consistency. In: B. L. Golden, E. A. Wasil & P. T. Harker, eds. *Analytic Hierarchy Process: Applications and Studies.* New York: Springer Verlag, pp. 68-81.

Harker, P. & Vargas, L., 1987. The Theory of Ratio Scale Estimation: Saaty's Analytic Hierarchy Process. *Management Science,* 33(11), pp. 1383-1403.

Herman, M. W. & Koczkodaj, W. W., 1995. A Monte Carlo Study of Pairwise Comparison. *Information Processing Letters*, pp. 223-239.

Hochbaum, D. S. & Levin, A., 2006. Methodologies for the Group Rankings Decision. *Management Science,* Volume 52, pp. 1394-1408.

Hwang, C. L. & Yoon, K., 1981. *Multiple Attribute Decision Making: Methods and Applications.* New York: Springer-Verlag.

Ishizaka, A., Balkenborg, D. & Kaplan, T., 2010. Influence of aggregation and measurement scale on ranking a compromise alternative in AHP. *Journal of the Operational Research Society,* 62(4), pp. 700-710.

Ishizaka, A. & Labib, A., 2011. Review of the Main Developments in the Analytic Hierarchy Process. *Expert Systems with Applications,* Volume 38, p. 14336–14345.

Ishizaka, A. & Lusti, M., 2006. How to Derive Priorities in AHP: A Comparative Study. *Central European Journal of Operations Research,* 14(4), pp. 387-400.

Işıklar, G. & Büyüközkan, G., 2007. Using a Multi-Criteria Decision Making Approach to Evaluate Mobile Phone Alternatives. *Computer Standards & Interfaces,* Volume 29, p. 265–274.

Ji, P. & Jiang, R., 2003. Scale Transitivity in the AHP. *The Journal of the Operational Research Society,* 54(8), pp. 896-905.

Johnson, C., Beine, W. & Wang, T., 1979. Right-Left Asymmetry in an Eigenvector Ranking Procedure. *Journal of Mathematical Psychology,* Volume 19, pp. 61-64.

Kendall, M. G., 1938. A New Measure of Rank Correlation. *Biometrika,* 30(1), pp. 81-93.

Kumar, V. N. & Ganesh, L., 1996. A Simulation-Based Evaluation of the Approximate and the Exact Eigenvector Methods Employed in AHP. *European Journal of Operational Research,* Volume 95, pp. 656-662.

Lane, E. F. & Verdini, W. A., 1989. A Consistency Test for AHP Decision Makers. *Decision Sciences,* Volume 20, pp. 575-590.

Lin, C., Koua, G. & Ergu, D., 2013. A Heuristic Approach for Deriving the Priority Vector in AHP. *Applied Mathematical Modelling,* 37(8), pp. 5828-5836.

Lootsma, F. A., 1989. Conflict Resolution via Pairwise Comparison of Concessions. *European Journal of Operational Research,* 40(1), pp. 109-116.

Lootsma, F. A., 1993. Scale Sensitivity in the Multiplicative AHP and SMART. *Journal of Multi-Criteria Decision Analysis,* Volume 2, pp. 87-110.

Lootsma, F. A., 1999. *Multi-Criteria Decision Analysis via Ratio and Difference Judgment.* Dordrecht: Kluwer Academic Publishers.

Macharis, C., Springael, J., De Brucker, K. & Verbeke, A., 2004. PROMETHEE and AHP: The Design of Operational Synergies in Multicriteria Analysis.: Strengthening PROMETHEE with Ideas of AHP. *European Journal of Operational Research,* 153(2), pp. 307-317.

Ma, D. & Zheng, X., 1991. *9/9-9/1 Scale Method of AHP.* Pittsburgh, s.n., pp. 197-202.

Mazurek, J. & Perzina, R., 2017. *On the Inconsistency of Pairwise Comparisons: An Experimental Study.* s.l.:s.n.

Miller, G., 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review,* 63(2), pp. 81-97.

Mu, E. & Pereyra-Rojas, M., 2017. An Introduction to the Analytic Hierarchy Process (AHP) Using Super Decisions V2. In: *Practical Decision Making.* Springer Online Documents: Springer International Publishing, pp. 7-22.

Murphy, C. K., 1993. Limits of the Analytical Hierarchy Process from Its Consistency Index. *European Journal of Operational Research,* Volume 65, pp. 138-139.

Noble, E. E. & Sanchez, P. P., 1993. A Note on The Information Content of a Consistent Pairwise Comparison Judgment Matrix of an AHP Decision Maker. *Theory and Decision,* 34(2), pp. 99-108.

Peláez, J. I. & Lamata, M. T., 2003. A New Measure of Consistency for Positive Reciprocal Matrices. *Computers and Mathematics with Applications,* Volume 46, pp. 1839-1845.

Pirdashti, M., Ghadi, A., Mohammadi, M. & Shojatalab, G., 2009. Multi-Criteria Decision-Making Selection Model with Application to Chemical Engineering Management Decisions. *World Academy of Science, Engineering and Technology International Journal of Chemical and Molecular Engineering,* 3(1), pp. 1-6.

Raharjo, H. & Endah, D., 2005. Evaluating Eelationship of Consistency Ratio and Number of Alternatives on Rank Reversal in the AHP. *Quality Engineering,* Volume 18, pp. 39-46.

Roy, B. & Bouyssou, D., 1993. *Aide Multicritére à la Décision: Méthodes et Cas.* Paris: Economica.

Russo, R. F. S. M. & Camanho, R., 2015. Criteria in AHP: A Systematic Review of Literature. *Procedia Computer Science,* Volume 55, pp. 1123-1132.

Saaty, R. W., 1987. The Analytic Hierarchy Process: What It Is and How It Is Used. *Mathematical Modelling,* 9(3-5), pp. 161-176.

Saaty, T. L., 1977. A Scaling Method for Priorities in Hierarchical Structures. *Journal of Mathematical Psychology,* 15(3), pp. 234-281.

Saaty, T. L., 1980. *The Analytic Hierarchy Process.* New York: McGraw Hill International.

Saaty, T. L., 1986. Axiomatic Foundation of the Analytic Hierarchy Process. *Management Science,* 32(7), pp. 841-855.

Saaty, T. L., 1990. Eigenvector and Logarithmic Least Squares. *European Journal of Operational Research,* Volume 48, pp. 156-160.

Saaty, T. L., 1994. How to Make a Decision: The Analytic Hierarchy Process. *Interfaces,* 24(6), pp. 19-43.

Saaty, T. L., 1996. *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy.* Pittsburgh: RWS Publications.

Saaty, T. L., 2001. The Seven Pillars of the Analytic Hierarchy Process: Models, Methods, Concepts & Applications of the Analytic Hierarchy Process. *International Series in Operations Research & Management Science,* Volume 34, pp. 27-46.

Saaty, T. L., 2003. Decision-Making with the AHP: Why is the Principal Eigenvector Necessary. *European Journal of Operational Research,* Volume 145, pp. 85-91.

Saaty, T. L., 2005. *Theory and Applications of the Analytic Network Process.* s.l.:RWS Publications.

Saaty, T. L. & Hu, G., 1998. Ranking by the Eigenvector Versus Other Methods in the Analytic Hierarchy Process. *Applied Mathematical Letters,* 11(4), pp. 121-125.

Saaty, T. L. & Sagir Ozdemir, M., 2003. Why the Magic Number Seven Plus or Minus Two. *Mathematical and Computer Modeling,* Volume 38, pp. 233-244.

Saaty, T. L. & Vargas, L. G., 2012. *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process.* 2 ed. New York: Springer US.

Safian, E. E. M. & Nawawi , A. H., 2011. *The Evolution of AHP as a Decision Making Tool in Property Sectors.* Bali/INDONESIA, IACSIT Press, pp. 28-31.

Sagir Ozdemir, M., 2005. Validity and Inconsistency in the Analytic Hierarchy Process. *Applied Mathematics and Computation,* 161(1), pp. 707-720.

Salo, A. A. & Hämäläinen, R. P., 1997. On the Measurement of Preferences in the Analytic Hierarchy Process. *Journal of Multi-Criteria Decision Analysis,* 6(6), pp. 309-319.

Schmidt, K. et al., 2015. Applying the Analytic Hierarchy Process in Healthcare Research. *BMC Medical Informatics & Decision Making,* pp. 15-112.

Simon, H. A., 1955. A Behavioral Model of Rational Choice. *Quarterly Journal of Economics,* 69(1), pp. 99-118.

Srdjevic, B., 2005. Combining Different Prioritization Methods in the Analytic Hierarchy Process Synthesis. *Computers & Operational Research,* Volume 32, pp. 1897-1919.

Stein, W. E. & Mizzi, P. J., 2007. The Harmonic Consistency Index for the Analytic Hierarchy Process. *European Journal of Operational Research,* Volume 177, pp. 488-497.

Stevens, S. S., 1957. On the Psychophysical Law. *Psychological Review,* Volume 64, pp. 153-181.

Stevens, S. S. & Galanter, E., 1964. Ratio Scales and Category Scales for a Dozen Perceptual Continua. *Journal of Experimental Psychology,* Volume 54, pp. 377-411.

Stirn, L. Z. & Grošelj, P., 2010. Multiple Critera Methods with Focus on Analytic Hierarchy Process and Group Decision Making. *Croatian Operational Research Review (CRORR),* Volume 1.

Takamura, Y. & Tone, K., 2003. A Comparative Site Evaluation Study for Relocating Japanese Government Agencies in Tokyo. *Socio-Economic Planning Sciences,* 27(2), pp. 85-102.

Takeda, E., Cogger, K. & Yu, P. L., 1987. Estimating Criterion Weights Using Eigenvectors: A Comparative Study. *European Journal of Operational Research,* Volume 29, pp. 360-369.

Tavana, M., 2003. Cross: A Multicriteria Group-Decision-Making Model for Evaluating and Prioritizing Advanced Technology Projects at NASA. *Interfaces - INFORMS,* 33(3), pp. 40-56.

Thurstone, L. L., 1927. A Law of Comparative Judgements. *Psychological Reviews,* 34(1), pp. 273-286.

Tomashevskii, I. L., 2014. *Geometric Mean Method for Judgement Matrices: Formulas for Errors,* Arkhangelsk: s.n.

Tomashevskii, I. L., 2015. Eigenvector ranking method as a measuring tool: Formulas for errors.. *European Journal of Operational Research,* Volume 240, pp. 774-780.

Triantaphyllou, E. & Mann, S. H., 1995. Using The Analytic Hierarchy Process For Decision Making In Engineering Applications: Some Challenges. *International Journal of Industrial Engineering: Applications and Practice,* 2(1), pp. 35-44.

Tummala, V. M. R. & Wan, Y.-w., 1994. On the Mean Random Inconsistency Index of Analytic Hierarchy Process (AHP). *Computers & Industrial Engineering,* 27(1), pp. 401-404.

Vaidya, O. S. & Kumar, S., 2006. Analytic Hierarchy Process: An Overview of Applications. *European Journal of Operational Research,* Volume 169, pp. 1-29.

von Winterfeldt, D. & Edwards, W., 1986. *Decision Analysis and Behavioral Research.* Cambridge: Cambridge University Press.

Zahedi, F., 1986. A Simulation Study of Estimation Methods in the Analytic Hierarchy Process. *Socio-Economic Planning Science,* Volume 20, pp. 347-354.

Zhang, Z., Liu, X. & Yang, S., 2009. A Note on the 1-9 Scale and Index Scale In AHP. *Communications in Computer and Information Science,* 35(1), pp. 630-634.

APPENDICES

A. SCALE BASED CONSISTENCY ASSESSMENTS

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

	9.000	**5.667**	**4.000**	**3.000**	**2.333**	**1.857**	**1.500**	**1.222**	**1.000**	**0.818**	**0.667**	**0.538**	**0.429**	**0.333**	**0.250**	**0.176**	**0.111**
9.000	81.000	51.000	36.000	27.000	21.000	16.714	13.500	11.000	9.000	7.364	6.000	4.846	3.857	3.000	2.250	1.588	1.000
5.667	51.000	32.111	22.667	17.000	13.222	10.524	8.500	6.926	5.667	4.636	3.778	3.051	2.429	1.889	1.417	1.000	0.630
4.000	36.000	22.667	16.000	12.000	9.333	7.429	6.000	4.889	4.000	3.273	2.667	2.154	1.714	1.333	1.000	0.706	0.444
3.000	27.000	17.000	12.000	9.000	7.000	5.571	4.500	3.667	3.000	2.455	2.000	1.615	1.286	1.000	0.750	0.529	0.333
2.333	21.000	13.222	9.333	7.000	5.444	4.333	3.500	2.852	2.333	1.909	1.556	1.256	1.000	0.778	0.583	0.412	0.259
1.857	16.714	10.524	7.429	5.571	4.333	3.449	2.786	2.270	1.857	1.519	1.238	1.000	0.796	0.619	0.464	0.328	0.206
1.500	13.500	8.500	6.000	4.500	3.500	2.786	2.250	1.833	1.500	1.227	1.000	0.808	0.643	0.500	0.375	0.265	0.167
1.222	11.000	6.926	4.889	3.667	2.852	2.270	1.833	1.494	1.222	1.000	0.815	0.658	0.524	0.407	0.306	0.216	0.136
1.000	9.000	5.667	4.000	3.000	2.333	1.857	1.500	1.222	1.000	0.818	0.667	0.538	0.429	0.333	0.250	0.176	0.111
0.818	7.364	4.636	3.273	2.455	1.909	1.519	1.227	1.000	0.818	0.669	0.545	0.441	0.351	0.273	0.205	0.144	0.091
0.667	6.000	3.778	2.667	2.000	1.556	1.238	1.000	0.815	0.667	0.545	0.444	0.359	0.286	0.222	0.167	0.118	0.074
0.538	4.846	3.051	2.154	1.615	1.256	1.000	0.808	0.658	0.538	0.441	0.359	0.290	0.231	0.179	0.135	0.095	0.060
0.429	3.857	2.429	1.714	1.286	1.000	0.796	0.643	0.524	0.429	0.351	0.286	0.231	0.184	0.143	0.107	0.076	0.048
0.333	3.000	1.889	1.333	1.000	0.778	0.619	0.500	0.407	0.333	0.273	0.222	0.179	0.143	0.111	0.083	0.059	0.037
0.250	2.250	1.417	1.000	0.750	0.583	0.464	0.375	0.306	0.250	0.205	0.167	0.135	0.107	0.083	0.063	0.044	0.028
0.176	1.588	1.000	0.706	0.529	0.412	0.328	0.265	0.216	0.176	0.144	0.118	0.095	0.076	0.059	0.044	0.031	0.020
0.111	1.000	0.630	0.444	0.333	0.259	0.206	0.167	0.136	0.111	0.091	0.074	0.060	0.048	0.037	0.028	0.020	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 the Exponential Scale

	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				

Table A 3 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 4 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Geometric Scale

	256.000	128.000	64.000	32.000	16.000	8.000	4.000	2.000	1.000	0.500	0.250	0.125	0.063	0.031	0.016	0.008	0.004
256.000	65536.00	32768.00	16384.00	8192.000	4096.000	2048.000	1024.000	512.000	256.000	128.000	64.000	32.000	16.000	8.000	4.000	2.000	1.000
128.000	32768.00	16384.00	8192.000	4096.000	2048.000	1024.000	512.000	256.000	128.000	64.000	32.000	16.000	8.000	4.000	2.000	1.000	0.500
64.000	16384.00	8192.000	4096.000	2048.000	1024.000	512.000	256.000	128.000	64.000	32.000	16.000	8.000	4.000	2.000	1.000	0.500	0.250
32.000	8192.000	4096.000	2048.000	1024.000	512.000	256.000	128.000	64.000	32.000	16.000	8.000	4.000	2.000	1.000	0.500	0.250	0.125
16.000	4096.000	2048.000	1024.000	512.000	256.000	128.000	64.000	32.000	16.000	8.000	4.000	2.000	1.000	0.500	0.250	0.125	0.063
8.000	2048.000	1024.000	512.000	256.000	128.000	64.000	32.000	16.000	8.000	4.000	2.000	1.000	0.500	0.250	0.125	0.063	0.031
4.000	1024.000	512.000	256.000	128.000	64.000	32.000	16.000	8.000	4.000	2.000	1.000	0.500	0.250	0.125	0.063	0.031	0.016
2.000	512.000	256.000	128.000	64.000	32.000	16.000	8.000	4.000	2.000	1.000	0.500	0.250	0.125	0.063	0.031	0.016	0.008
1.000	256.000	128.000	64.000	32.000	16.000	8.000	4.000	2.000	1.000	0.500	0.250	0.125	0.063	0.031	0.016	0.008	0.004
0.500	128.000	64.000	32.000	16.000	8.000	4.000	2.000	1.000	0.500	0.250	0.125	0.063	0.031	0.016	0.008	0.004	0.002
0.250	64.000	32.000	16.000	8.000	4.000	2.000	1.000	0.500	0.250	0.125	0.063	0.031	0.016	0.008	0.004	0.002	0.001
0.125	32.000	16.000	8.000	4.000	2.000	1.000	0.500	0.250	0.125	0.063	0.031	0.016	0.008	0.004	0.002	0.001	0.000
0.063	16.000	8.000	4.000	2.000	1.000	0.500	0.250	0.125	0.063	0.031	0.016	0.008	0.004	0.002	0.001	0.000	0.000
0.031	8.000	4.000	2.000	1.000	0.500	0.250	0.125	0.063	0.031	0.016	0.008	0.004	0.002	0.001	0.000	0.000	0.000
0.016	4.000	2.000	1.000	0.500	0.250	0.125	0.063	0.031	0.016	0.008	0.004	0.002	0.001	0.000	0.000	0.000	0.000
0.008	2.000	1.000	0.500	0.250	0.125	0.063	0.031	0.016	0.008	0.004	0.002	0.001	0.000	0.000	0.000	0.000	0.000
0.004	1.000	0.500	0.250	0.125	0.063	0.031	0.016	0.008	0.004	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Color Codes	Within the limits and defined by the scale						Within the limits of the scale					Outside the limits of the scale					

Table A 5 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Inverse Linear Scale

	9.000	**4.500**	**3.000**	**2.250**	**1.800**	**1.500**	**1.286**	**1.125**	**1.000**	**0.889**	**0.778**	**0.667**	**0.556**	**0.444**	**0.333**	**0.222**	**0.111**
9.000	81.000	40.500	27.000	20.250	16.200	13.500	11.571	10.125	9.000	8.000	7.000	6.000	5.000	4.000	3.000	2.000	1.000
4.500	40.500	20.250	13.500	10.125	8.100	6.750	5.786	5.063	4.500	4.000	3.500	3.000	2.500	2.000	1.500	1.000	0.500
3.000	27.000	13.500	9.000	6.750	5.400	4.500	3.857	3.375	3.000	2.667	2.333	2.000	1.667	1.333	1.000	0.667	0.333
2.250	20.250	10.125	6.750	5.063	4.050	3.375	2.893	2.531	2.250	2.000	1.750	1.500	1.250	1.000	0.750	0.500	0.250
1.800	16.200	8.100	5.400	4.050	3.240	2.700	2.314	2.025	1.800	1.600	1.400	1.200	1.000	0.800	0.600	0.400	0.200
1.500	13.500	6.750	4.500	3.375	2.700	2.250	1.929	1.688	1.500	1.333	1.167	1.000	0.833	0.667	0.500	0.333	0.167
1.286	11.571	5.786	3.857	2.893	2.314	1.929	1.653	1.446	1.286	1.143	1.000	0.857	0.714	0.571	0.429	0.286	0.143
1.125	10.125	5.063	3.375	2.531	2.025	1.688	1.446	1.266	1.125	1.000	0.875	0.750	0.625	0.500	0.375	0.250	0.125
1.000	9.000	4.500	3.000	2.250	1.800	1.500	1.286	1.125	1.000	0.889	0.778	0.667	0.556	0.444	0.333	0.222	0.111
0.889	8.000	4.000	2.667	2.000	1.600	1.333	1.143	1.000	0.889	0.790	0.691	0.593	0.494	0.395	0.296	0.198	0.099
0.778	7.000	3.500	2.333	1.750	1.400	1.167	1.000	0.875	0.778	0.691	0.605	0.519	0.432	0.346	0.259	0.173	0.086
0.667	6.000	3.000	2.000	1.500	1.200	1.000	0.857	0.750	0.667	0.593	0.519	0.444	0.370	0.296	0.222	0.148	0.074
0.556	5.000	2.500	1.667	1.250	1.000	0.833	0.714	0.625	0.556	0.494	0.432	0.370	0.309	0.247	0.185	0.123	0.062
0.444	4.000	2.000	1.333	1.000	0.800	0.667	0.571	0.500	0.444	0.395	0.346	0.296	0.247	0.198	0.148	0.099	0.049
0.333	3.000	1.500	1.000	0.750	0.600	0.500	0.429	0.375	0.333	0.296	0.259	0.222	0.185	0.148	0.111	0.074	0.037
0.222	2.000	1.000	0.667	0.500	0.400	0.333	0.286	0.250	0.222	0.198	0.173	0.148	0.123	0.099	0.074	0.049	0.025
0.111	1.000	0.500	0.333	0.250	0.200	0.167	0.143	0.125	0.111	0.099	0.086	0.074	0.062	0.049	0.037	0.025	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 6 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Logarithmic Scale

	3.322	**3.170**	**3.000**	**2.807**	**2.585**	**2.322**	**2.000**	**1.585**	**1.000**	**0.631**	**0.500**	**0.431**	**0.387**	**0.356**	**0.333**	**0.315**	**0.301**
3.322	11.035	10.530	9.966	9.326	8.587	7.713	6.644	5.265	3.322	2.096	1.661	1.431	1.285	1.183	1.107	1.048	1.000
3.170	10.530	10.048	9.510	8.899	8.194	7.360	6.340	5.024	3.170	2.000	1.585	1.365	1.226	1.129	1.057	1.000	0.954
3.000	9.966	9.510	9.000	8.422	7.755	6.966	6.000	4.755	3.000	1.893	1.500	1.292	1.161	1.069	1.000	0.946	0.903
2.807	9.326	8.899	8.422	7.881	7.257	6.518	5.615	4.450	2.807	1.771	1.404	1.209	1.086	1.000	0.936	0.886	0.845
2.585	8.587	8.194	7.755	7.257	6.682	6.002	5.170	4.097	2.585	1.631	1.292	1.113	1.000	0.921	0.862	0.815	0.778
2.322	7.713	7.360	6.966	6.518	6.002	5.391	4.644	3.680	2.322	1.465	1.161	1.000	0.898	0.827	0.774	0.732	0.699
2.000	6.644	6.340	6.000	5.615	5.170	4.644	4.000	3.170	2.000	1.262	1.000	0.861	0.774	0.712	0.667	0.631	0.602
1.585	5.265	5.024	4.755	4.450	4.097	3.680	3.170	2.512	1.585	1.000	0.792	0.683	0.613	0.565	0.528	0.500	0.477
1.000	3.322	3.170	3.000	2.807	2.585	2.322	2.000	1.585	1.000	0.631	0.500	0.431	0.387	0.356	0.333	0.315	0.301
0.631	2.096	2.000	1.893	1.771	1.631	1.465	1.262	1.000	0.631	0.398	0.315	0.272	0.244	0.225	0.210	0.199	0.190
0.500	1.661	1.585	1.500	1.404	1.292	1.161	1.000	0.792	0.500	0.315	0.250	0.215	0.193	0.178	0.167	0.158	0.151
0.431	1.431	1.365	1.292	1.209	1.113	1.000	0.861	0.683	0.431	0.272	0.215	0.185	0.167	0.153	0.144	0.136	0.130
0.387	1.285	1.226	1.161	1.086	1.000	0.898	0.774	0.613	0.387	0.244	0.193	0.167	0.150	0.138	0.129	0.122	0.116
0.356	1.183	1.129	1.069	1.000	0.921	0.827	0.712	0.565	0.356	0.225	0.178	0.153	0.138	0.127	0.119	0.112	0.107
0.333	1.107	1.057	1.000	0.936	0.862	0.774	0.667	0.528	0.333	0.210	0.167	0.144	0.129	0.119	0.111	0.105	0.100
0.315	1.048	1.000	0.946	0.886	0.815	0.732	0.631	0.500	0.315	0.199	0.158	0.136	0.122	0.112	0.105	0.100	0.095
0.301	1.000	0.954	0.903	0.845	0.778	0.699	0.602	0.477	0.301	0.190	0.151	0.130	0.116	0.107	0.100	0.095	0.091
Color Codes	Within the limits and defined by the scale							Within the limits of the scale					Outside the limits of the scale				

Table A 7 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Power Scale

	81.000	**64.000**	**49.000**	**36.000**	**25.000**	**16.000**	**9.000**	**4.000**	**1.000**	**0.250**	**0.111**	**0.063**	**0.040**	**0.028**	**0.020**	**0.016**	**0.012**
81.000	6561.000	5184.000	3969.000	2916.000	2025.000	1296.000	729.000	324.000	81.000	20.250	9.000	5.063	3.240	2.250	1.653	1.266	1.000
64.000	5184.000	4096.000	3136.000	2304.000	1600.000	1024.000	576.000	256.000	64.000	16.000	7.111	4.000	2.560	1.778	1.306	1.000	0.790
49.000	3969.000	3136.000	2401.000	1764.000	1225.000	784.000	441.000	196.000	49.000	12.250	5.444	3.063	1.960	1.361	1.000	0.766	0.605
36.000	2916.000	2304.000	1764.000	1296.000	900.000	576.000	324.000	144.000	36.000	9.000	4.000	2.250	1.440	1.000	0.735	0.563	0.444
25.000	2025.000	1600.000	1225.000	900.000	625.000	400.000	225.000	100.000	25.000	6.250	2.778	1.563	1.000	0.694	0.510	0.391	0.309
16.000	1296.000	1024.000	784.000	576.000	400.000	256.000	144.000	64.000	16.000	4.000	1.778	1.000	0.640	0.444	0.327	0.250	0.198
9.000	729.000	576.000	441.000	324.000	225.000	144.000	81.000	36.000	9.000	2.250	1.000	0.563	0.360	0.250	0.184	0.141	0.111
4.000	324.000	256.000	196.000	144.000	100.000	64.000	36.000	16.000	4.000	1.000	0.444	0.250	0.160	0.111	0.082	0.063	0.049
1.000	81.000	64.000	49.000	36.000	25.000	16.000	9.000	4.000	1.000	0.250	0.111	0.063	0.040	0.028	0.020	0.016	0.012
0.250	20.250	16.000	12.250	9.000	6.250	4.000	2.250	1.000	0.250	0.063	0.028	0.016	0.010	0.007	0.005	0.004	0.003
0.111	9.000	7.111	5.444	4.000	2.778	1.778	1.000	0.444	0.111	0.028	0.012	0.007	0.004	0.003	0.002	0.002	0.001
0.063	5.063	4.000	3.063	2.250	1.563	1.000	0.563	0.250	0.063	0.016	0.007	0.004	0.003	0.002	0.001	0.001	0.001
0.040	3.240	2.560	1.960	1.440	1.000	0.640	0.360	0.160	0.040	0.010	0.004	0.003	0.002	0.001	0.001	0.001	0.000
0.028	2.250	1.778	1.361	1.000	0.694	0.444	0.250	0.111	0.028	0.007	0.003	0.002	0.001	0.001	0.001	0.000	0.000
0.020	1.653	1.306	1.000	0.735	0.510	0.327	0.184	0.082	0.020	0.005	0.002	0.001	0.001	0.001	0.000	0.000	0.000
0.016	1.266	1.000	0.766	0.563	0.391	0.250	0.141	0.063	0.016	0.004	0.002	0.001	0.001	0.000	0.000	0.000	0.000
0.012	1.000	0.790	0.605	0.444	0.309	0.198	0.111	0.049	0.012	0.003	0.001	0.001	0.000	0.000	0.000	0.000	0.000
Color Codes	Within the limits and defined by the scale								Within the limits of the scale				Outside the limits of the scale				

Table A 8 Possible (a_{ij}, a_{jk}) pairs and the resultant a_{ik} values in the Root Square Scale

	3.000	2.828	2.646	2.449	2.236	2.000	1.732	1.414	1.000	0.707	0.577	0.500	0.447	0.408	0.378	0.354	0.333
3.000	9.000	8.485	7.937	7.348	6.708	6.000	5.196	4.243	3.000	2.121	1.732	1.500	1.342	1.225	1.134	1.061	1.000
2.828	8.485	8.000	7.483	6.928	6.325	5.657	4.899	4.000	2.828	2.000	1.633	1.414	1.265	1.155	1.069	1.000	0.943
2.646	7.937	7.483	7.000	6.481	5.916	5.292	4.583	3.742	2.646	1.871	1.528	1.323	1.183	1.080	1.000	0.935	0.882
2.449	7.348	6.928	6.481	6.000	5.477	4.899	4.243	3.464	2.449	1.732	1.414	1.225	1.095	1.000	0.926	0.866	0.816
2.236	6.708	6.325	5.916	5.477	5.000	4.472	3.873	3.162	2.236	1.581	1.291	1.118	1.000	0.913	0.845	0.791	0.745
2.000	6.000	5.657	5.292	4.899	4.472	4.000	3.464	2.828	2.000	1.414	1.155	1.000	0.894	0.816	0.756	0.707	0.667
1.732	5.196	4.899	4.583	4.243	3.873	3.464	3.000	2.449	1.732	1.225	1.000	0.866	0.775	0.707	0.655	0.612	0.577
1.414	4.243	4.000	3.742	3.464	3.162	2.828	2.449	2.000	1.414	1.000	0.816	0.707	0.632	0.577	0.535	0.500	0.471
1.000	3.000	2.828	2.646	2.449	2.236	2.000	1.732	1.414	1.000	0.707	0.577	0.500	0.447	0.408	0.378	0.354	0.333
0.707	2.121	2.000	1.871	1.732	1.581	1.414	1.225	1.000	0.707	0.500	0.408	0.354	0.316	0.289	0.267	0.250	0.236
0.577	1.732	1.633	1.528	1.414	1.291	1.155	1.000	0.816	0.577	0.408	0.333	0.289	0.258	0.236	0.218	0.204	0.192
0.500	1.500	1.414	1.323	1.225	1.118	1.000	0.866	0.707	0.500	0.354	0.289	0.250	0.224	0.204	0.189	0.177	0.167
0.447	1.342	1.265	1.183	1.095	1.000	0.894	0.775	0.632	0.447	0.316	0.258	0.224	0.200	0.183	0.169	0.158	0.149
0.408	1.225	1.155	1.080	1.000	0.913	0.816	0.707	0.577	0.408	0.289	0.236	0.204	0.183	0.167	0.154	0.144	0.136
0.378	1.134	1.069	1.000	0.926	0.845	0.756	0.655	0.535	0.378	0.267	0.218	0.189	0.169	0.154	0.143	0.134	0.126
0.354	1.061	1.000	0.935	0.866	0.791	0.707	0.612	0.500	0.354	0.250	0.204	0.177	0.158	0.144	0.134	0.125	0.118
0.333	1.000	0.943	0.882	0.816	0.745	0.667	0.577	0.471	0.333	0.236	0.192	0.167	0.149	0.136	0.126	0.118	0.111
Color Codes	Within the limits and defined by the scale							Within the limits of the scale					Outside the limits of the scale				

B. SIMULATION 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	Number of Deviations (Different Best)
3x3	0.150	62.50%	37.40%	0.10%	29.80%	16.60%	53.60%	96.40%	96.30%	0.9508	0.1736	36
	0.100	62.40%	37.60%	0.00%	34.10%	21.20%	44.70%	96.70%	97.20%	0.9576	0.1653	33
	0.050	61.80%	38.20%	0.00%	54.80%	23.30%	21.90%	100.00%	99.70%	0.9974	0.0378	0
4x4	0.150	63.40%	36.60%	0.00%	30.20%	20.80%	49.00%	91.50%	90.60%	0.9107	0.1632	85
	0.100	70.30%	29.70%	0.00%	45.20%	18.40%	36.40%	90.90%	93.00%	0.9205	0.1547	91
	0.050	69.80%	30.20%	0.00%	58.50%	24.40%	17.10%	93.90%	95.20%	0.9442	0.1360	61
5x5	0.150	70.20%	29.80%	0.00%	46.70%	20.00%	33.30%	88.20%	88.00%	0.9028	0.1383	118
	0.100	70.20%	29.80%	0.00%	53.60%	21.90%	24.50%	90.80%	89.30%	0.9226	0.1232	92
	0.050	76.80%	23.20%	0.00%	69.90%	20.40%	9.70%	92.90%	92.60%	0.9437	0.0992	71
6x6	0.150	75.50%	24.50%	0.00%	62.80%	17.70%	19.50%	88.10%	89.70%	0.9056	0.1076	119
	0.100	82.60%	17.40%	0.00%	73.10%	13.00%	13.90%	87.90%	89.70%	0.9126	0.1055	121
	0.050	85.50%	14.50%	0.00%	84.00%	12.20%	3.80%	90.70%	92.30%	0.9392	0.0881	93
7x7	0.150	84.20%	15.80%	0.00%	75.30%	11.60%	13.10%	86.70%	87.00%	0.9029	0.0904	133
	0.100	90.70%	9.30%	0.00%	85.10%	7.30%	7.60%	86.50%	88.30%	0.9161	0.0896	135

Table B 2 Performance measures for Exponential 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	Number of Deviations (Different Best)
3x3	0.150	67.00%	32.90%	0.10%	25.00%	15.40%	59.60%	96.40%	97.30%	0.9555	0.1643	36
	0.100	68.90%	31.10%	0.00%	41.00%	11.90%	47.10%	97.40%	97.00%	0.9608	0.1540	26
	0.050	65.50%	34.40%	0.10%	54.70%	17.40%	27.90%	100.00%	99.80%	0.9961	0.0367	0
4x4	0.150	85.20%	14.80%	0.00%	39.90%	4.00%	56.10%	91.90%	92.30%	0.9182	0.1585	81
	0.100	83.10%	16.90%	0.00%	50.30%	5.60%	44.10%	93.10%	95.20%	0.9407	0.1341	69
	0.050	83.00%	17.00%	0.00%	62.90%	8.10%	29.00%	94.80%	95.80%	0.9517	0.1205	52
5x5	0.150	95.10%	4.90%	0.00%	66.90%	1.40%	31.70%	91.50%	91.60%	0.9207	0.1229	85
	0.100	95.60%	4.40%	0.00%	76.00%	1.20%	22.80%	90.50%	90.90%	0.9263	0.1164	95
	0.050	94.10%	5.90%	0.00%	83.50%	3.10%	13.40%	92.60%	94.40%	0.9431	0.1019	74
6x6	0.150	98.50%	1.50%	0.00%	83.20%	0.70%	16.10%	87.10%	91.30%	0.9164	0.1009	129
	0.100	98.70%	1.30%	0.00%	90.00%	0.30%	9.70%	88.40%	90.80%	0.9212	0.0991	116
	0.050	99.10%	0.90%	0.00%	96.50%	0.40%	3.10%	91.00%	90.50%	0.9337	0.0885	90
7x7	0.150	99.40%	0.60%	0.00%	94.00%	0.10%	5.90%	86.80%	90.50%	0.9136	0.0919	132
	0.100	99.90%	0.10%	0.00%	97.70%	0.00%	2.30%	90.40%	90.40%	0.9282	0.0826	96

Table B 3 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	Number of Deviations (Different Best)
3x3	0.150	76.70%	23.30%	0.00%	35.30%	12.70%	52.00%	97.90%	97.30%	0.9665	0.1432	21
	0.100	72.60%	27.40%	0.00%	36.00%	19.50%	44.50%	99.20%	98.50%	0.9821	0.1019	8
	0.050	69.80%	30.20%	0.00%	49.60%	23.70%	26.70%	100.00%	99.70%	0.9969	0.0391	0
4x4	0.150	97.00%	3.00%	0.00%	65.10%	0.10%	34.80%	90.20%	92.90%	0.9125	0.1645	98
	0.100	96.30%	3.70%	0.00%	69.30%	0.70%	30.00%	91.10%	93.10%	0.9207	0.1548	89
	0.050	94.50%	5.50%	0.00%	80.90%	1.40%	17.70%	94.60%	95.40%	0.9467	0.1324	54
5x5	0.150	99.90%	0.10%	0.00%	85.10%	0.00%	14.90%	87.50%	90.50%	0.8964	0.1400	125
	0.100	100.00%	0.00%	0.00%	91.60%	0.00%	8.40%	88.40%	89.80%	0.9048	0.1274	116
	0.050	99.10%	0.90%	0.00%	96.60%	0.10%	3.30%	89.80%	91.40%	0.9230	0.1133	102
6x6	0.150	100.00%	0.00%	0.00%	96.30%	0.00%	3.70%	84.30%	89.70%	0.8835	0.1229	157
	0.100	100.00%	0.00%	0.00%	97.90%	0.00%	2.10%	87.50%	87.90%	0.8989	0.1095	125
	0.050	100.00%	0.00%	0.00%	99.90%	0.00%	0.10%	88.10%	88.30%	0.9137	0.1013	119
7x7	0.150	100.00%	0.00%	0.00%	99.40%	0.00%	0.60%	82.60%	86.80%	0.8765	0.1112	174
	0.100	100.00%	0.00%	0.00%	99.80%	0.00%	0.20%	83.00%	88.30%	0.8959	0.0989	170

Table B 4 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	Number of Deviations (Different Best)
3x3	0.150	62.30%	37.60%	0.10%	33.90%	21.40%	44.70%	95.60%	95.40%	0.9385	0.1911	44
	0.100	63.50%	36.50%	0.00%	41.60%	22.20%	36.20%	95.90%	96.30%	0.9457	0.1857	41
	0.050	63.10%	36.80%	0.10%	55.90%	24.90%	19.20%	97.90%	98.10%	0.9706	0.1405	21
4x4	0.150	67.10%	32.90%	0.00%	46.60%	24.50%	28.90%	86.20%	87.20%	0.8715	0.1911	138
	0.100	75.00%	25.00%	0.00%	57.00%	18.80%	24.20%	90.20%	89.50%	0.8955	0.1720	98
	0.050	71.30%	28.70%	0.00%	63.10%	23.60%	13.30%	93.30%	93.50%	0.9342	0.1475	67
5x5	0.150	76.40%	23.60%	0.00%	65.50%	18.50%	16.00%	83.90%	85.40%	0.8525	0.1728	161
	0.100	81.10%	18.90%	0.00%	74.10%	16.10%	9.80%	82.80%	88.70%	0.8822	0.1426	172
	0.050	85.10%	14.90%	0.00%	82.70%	13.60%	3.70%	86.30%	87.70%	0.9025	0.1277	137
6x6	0.150	87.00%	13.00%	0.00%	81.90%	11.00%	7.10%	83.20%	81.80%	0.8557	0.1335	168
	0.100	90.40%	9.60%	0.00%	88.00%	8.70%	3.30%	83.90%	85.30%	0.8740	0.1226	161
	0.050	94.10%	5.90%	0.00%	93.80%	5.40%	0.80%	86.60%	87.60%	0.9031	0.1134	134
7x7	0.150	95.70%	4.30%	0.00%	94.10%	3.70%	2.20%	80.10%	81.00%	0.8587	0.1096	199
	0.100	97.20%	2.80%	0.00%	96.80%	2.60%	0.60%	84.60%	83.80%	0.8726	0.1102	154

Table B 5 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	Number of Deviations (Different Best)
3x3	0.150	27.70%	72.20%	0.10%	4.30%	14.70%	81.00%	99.40%	100.00%	0.9951	0.0530	6
	0.100	34.50%	65.50%	0.00%	6.50%	20.90%	72.60%	99.60%	100.00%	0.9953	0.0461	4
	0.050	44.20%	55.80%	0.00%	13.70%	17.70%	68.60%	99.40%	99.70%	0.9912	0.0666	6
4x4	0.150	3.10%	96.90%	0.00%	0.00%	32.10%	67.90%	99.60%	98.90%	0.9912	0.0531	4
	0.100	5.10%	94.90%	0.00%	0.20%	42.90%	56.90%	99.30%	99.30%	0.9928	0.0471	7
	0.050	11.50%	88.50%	0.00%	0.50%	46.20%	53.30%	99.40%	99.90%	0.9948	0.0397	6
5x5	0.150	0.30%	99.70%	0.00%	0.00%	53.20%	46.80%	98.20%	97.60%	0.9832	0.0552	18
	0.100	0.60%	99.40%	0.00%	0.00%	60.40%	39.60%	98.10%	98.50%	0.9875	0.0491	19
	0.050	1.30%	98.70%	0.00%	0.10%	69.40%	30.50%	99.00%	99.10%	0.9927	0.0380	10
6x6	0.150	0.00%	100.00%	0.00%	0.00%	67.50%	32.50%	96.80%	96.90%	0.9797	0.0525	32
	0.100	0.00%	100.00%	0.00%	0.00%	78.00%	22.00%	97.70%	97.70%	0.9845	0.0467	23
	0.050	0.40%	99.60%	0.00%	0.10%	84.30%	15.60%	98.20%	97.60%	0.9886	0.0395	18
7x7	0.150	0.00%	100.00%	0.00%	0.00%	83.80%	16.20%	96.40%	97.50%	0.9806	0.0414	36
	0.100	0.00%	100.00%	0.00%	0.00%	90.80%	9.20%	98.80%	97.00%	0.9842	0.0372	12

Table B 6 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	Number of Deviations (Different Best)
3x3	0.150	98.60%	1.30%	0.10%	24.20%	0.80%	75.00%	99.40%	99.80%	0.9934	0.0612	6
	0.100	98.30%	1.70%	0.00%	27.30%	0.40%	72.30%	99.60%	99.90%	0.9963	0.0477	4
	0.050	96.80%	3.00%	0.20%	39.50%	0.90%	59.60%	99.30%	99.20%	0.9891	0.0820	7
4x4	0.150	100.00%	0.00%	0.00%	57.70%	0.00%	42.30%	99.50%	98.70%	0.9912	0.0532	5
	0.100	99.90%	0.00%	0.10%	58.80%	0.00%	41.20%	98.90%	98.90%	0.9867	0.0664	11
	0.050	100.00%	0.00%	0.00%	61.50%	0.00%	38.50%	99.50%	99.00%	0.9943	0.0432	5
5x5	0.150	100.00%	0.00%	0.00%	74.70%	0.00%	25.30%	98.30%	97.50%	0.9785	0.0668	17
	0.100	100.00%	0.00%	0.00%	83.30%	0.00%	16.70%	98.30%	97.80%	0.9814	0.0581	17
	0.050	100.00%	0.00%	0.00%	87.00%	0.00%	13.00%	98.30%	97.50%	0.9862	0.0523	17
6x6	0.150	100.00%	0.00%	0.00%	88.70%	0.00%	11.30%	97.30%	95.80%	0.9704	0.0649	27
	0.100	100.00%	0.00%	0.00%	94.90%	0.00%	5.10%	97.10%	96.60%	0.9748	0.0577	29
	0.050	100.00%	0.00%	0.00%	96.50%	0.00%	3.50%	97.30%	98.40%	0.9862	0.0406	27
7x7	0.150	100.00%	0.00%	0.00%	97.30%	0.00%	2.70%	94.30%	95.70%	0.9665	0.0562	57
	0.100	100.00%	0.00%	0.00%	98.30%	0.00%	1.70%	94.80%	96.70%	0.9738	0.0483	52

Table B 7 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	Number of Deviations (Different Best)
3x3	0.150	1.40%	98.30%	0.30%	3.10%	3.50%	93.40%	99.20%	99.70%	0.9908	0.0717	8
	0.100	1.00%	98.70%	0.30%	2.90%	8.60%	88.50%	99.50%	99.30%	0.9902	0.0747	5
	0.050	1.10%	98.60%	0.30%	8.90%	0.70%	90.40%	99.00%	99.40%	0.9844	0.0882	10
4x4	0.150	0.00%	100.00%	0.00%	0.00%	24.90%	75.10%	99.50%	99.50%	0.9954	0.0380	5
	0.100	0.00%	100.00%	0.00%	0.00%	28.00%	72.00%	99.60%	99.50%	0.9957	0.0367	4
	0.050	0.00%	100.00%	0.00%	0.00%	27.90%	72.10%	99.80%	99.70%	0.9961	0.0313	2
5x5	0.150	0.00%	100.00%	0.00%	0.00%	45.40%	54.60%	99.10%	99.10%	0.9901	0.0432	9
	0.100	0.00%	100.00%	0.00%	0.00%	51.50%	48.50%	99.00%	98.30%	0.9910	0.0424	10
	0.050	0.00%	100.00%	0.00%	0.00%	52.90%	47.10%	99.60%	99.10%	0.9955	0.0294	4
6x6	0.150	0.00%	100.00%	0.00%	0.00%	65.30%	34.70%	97.40%	98.50%	0.9848	0.0460	26
	0.100	0.00%	100.00%	0.00%	0.00%	67.50%	32.50%	98.40%	98.80%	0.9897	0.0370	16
	0.050	0.00%	100.00%	0.00%	0.00%	70.90%	29.10%	99.00%	99.20%	0.9949	0.0262	10
7x7	0.150	0.00%	100.00%	0.00%	0.00%	76.10%	23.90%	97.60%	95.80%	0.9855	0.0365	24
	0.100	0.00%	100.00%	0.00%	0.00%	82.60%	17.40%	97.00%	98.30%	0.9890	0.0321	30

C. TURKISH SUMMARY/TÜRKÇE ÖZET

Günlük hayatta karşılaşılan sorunlar nadiren sadece bir değişkene bağlıdır. Aksine, bu sorunlar genelde birden çok seçenek, değerlendirme ölçütü ve paydaş içeren karmaşık sorunlardır. Yönetsel kararlar için ise bu karmaşıklık çok daha fazladır. Yönetsel karar vericilerin, karar verme sürecinde bütün (ya da en önemli) seçenekleri, değerlendirme ölçütlerini ve paydaşları aynı anda göz önünde bulundurmaları gerekir. Bu karmaşıklık ise iyi yapılandırılmış bir çok-kriterli karar verme yöntemine olan kaçınılmaz ihtiyacı ortaya serer.

Thomas L. Saaty (1980) tarafından ortaya atılmış olan Analitik Hiyerarşi Süreci (AHS) özel sektör, kamu ve savunma sanayiinde karşılaşılan önemli karar verme sorunlarında, çeşitli seviyelerde yöneticiler tarafından sıklıkla kullanılan bir çok-kriterli karar verme yöntemidir. Bu yöntem proje yönetimi, çevre yönetimi ve hatta kişisel kararlar için bile kullanılmaktadır. AHS ile bir grup seçenek arasından en iyisi seçilebilir; çeşitli seçenekler değerlendirilebilir ve performansları ölçülebilir; kaynak paylaştırma çalışmaları ve ileriye dönük tahmin gibi birçok önemli çalışmaya temel oluşturulabilir.

Yönetsel karar vericiler sıklıkla nesnel/sayısal veriyle birlikte bir paydaşın bir seçeneğe olan muhtemel tepkisi gibi öznel/sayısal olmayan verileri de kararlarında gözetmek zorunda kalırlar. AHS, karar vericilere doğrudan ölçülebilir nesnel veri ile doğrudan ölçülemez öznel veriyi sistematik olarak birleştirerek karar verme sürecinde birlikte değerlendirme şansı sunar.

AHS'nin yaygın kullanımının ve gücünün altındaki temel sebep, karmaşık olan geniş kapsamlı bir sorunu sistematik olarak daha küçük ve yönetilebilir parçalara bölmesi; daha sonrasında ise bütün sorunla aynı anda uğraşmak yerine daha küçük sorunlarla tek tek ilgilenilmesini sağlamasıdır. Yöntemin bir başka faydası ise değerlendirilen elemanların hepsini aynı anda düşünmek yerine bütün elemanları ikili kombinasyonlar halinde karşılaştırmaya olanak sağlamasıdır. İkili karşılaştırma yöntemi, aynı anda yalnızca iki elemanı karşılaştırmanın daha kolay olması ve daha isabetli sonuçlar vermesi sebebiyle psikologlar tarafından sıklıkla önerilmektedir. AHS'nin bir diğer güçlü yanı da karar verme süreci sonrasında karar vericinin değerlendirmelerindeki tutarlılığının sistematik şekilde değerlendirilmesine olanak vermesidir. AHS yöntemi, karar verme uygulamalarında sıklıkla tercih edilmesine katkıda bulunan iyi tanımlanmış bir tutarlılık kontrol sürecine sahiptir.

AHS yöntemi genel anlamda beş ana adımdan oluşmaktadır:

- Karar verme sorununun tanımlanması ve hiyerarşik olarak gösterilmesi
- İkili karşılaştırmalar
- Ağırlık belirleme
- Tutarlılık değerlendirmesi
- Ağırlıkların bütünleştirilmesi

AHS'nin uygulanabilmesi için karar verme sorununun net bir şekilde tanımlanmış ve süreç sonucunda neye ulaşılmak istendiğinin belirlenmiş olması gerekmektedir. Ulaşılmak istenen nokta "ana hedef" olarak tanımlanır. Daha sonrasında karmaşık karar verme sorunu, ana hedefe ulaşmak için en genel anlamda sağlanması gereken ana ölçütleri belirtecek şekilde bölümlere ayrılır ve bir hiyerarşi oluşturulur. Bu hiyerarşinin en üst noktasında ise ana hedef bulunur. Eğer belirlenen ana ölçütler ikili karşılaştırma yapmak için fazlasıyla genel kalıyorsa, bu ana ölçütler tekrar alt ölçütlere bölünebilir. Bu bölme işlemi ihtiyaç duyulduğu kadar tekrar edilebilir. Fakat, bölme işlemini yaparken göz önünde bulundurulması gereken en önemli nokta, oluşturulan alt ölçütlerin birbiriyle karşılaştırılabilecek ortak noktalarının kaybedilmemesidir. Eğer bir karar verme sorunu en alt seviyede bulunan ve aynı üst ölçüte bağlı olan alt ölçütlerin birbirleriyle karşılaştırılacak ortak noktası bulunmayacak kadar detaylandırıldıysa, bu detaylandırma sorunu basit bölümlere ayırmaktan ziyade daha karmaşık hale getirecektir. Ana hedefin altında bulunan bütün ölçüt seviyeleri ve hiyerarşisi belirlendikten sonra, elde bulunan seçenekler en alt seviyede bulunan her ölçütün altına yerleştirilir.

Sorunun hiyerarşik olarak ayrılması ve düzenlenmesinden sonra ikili karşılaştırma adımına geçilir. İkili karşılaştırmalarda sıklıkla kullanılan bir yönteme göre, en alt seviyede bulunan bütün ölçütler bir üst seviyede bağlı bulundukları ölçütlere göre ikili olarak karşılaştırılırlar. En alt seviyedeki bütün ölçütler tamamlandıktan sonra aynı işlem bir üst seviyedeki ölçütler için tekrarlanır. Bu adım ana hedefin hemen altında bulunan ana ölçütler de ikili karşılaştırılıncaya kadar tekrar edilir. Daha sonra bütün seçenekler, hiyerarşinin en alt seviyesinde bulunan bütün ölçütlerdeki performanslarına göre ikili olarak karşılaştırılır. Bu ikili karşılaştırmalar sözlü bir ölçek kullanılarak yapılır. Sonrasında ise bu sözel karşılaştırmalar, sayısal bir ölçekle birebir eşleştirme yöntemiyle oransal değerlendirmelere dönüştürülür. AHS yönteminin önerildiği tarihten bugüne birçok sayısal ikili karşılaştırma ölçeği (Dengeli, Temel, Geometrik, Ters Lineer, Logaritmik, Kare ve Karekök) önerilmiş olup hangi ölçeğin daha iyi olduğu tartışması günümüzde bile devam etmektedir. Yine de, öne sürüldüğü günden bu yana, basitliği ve anlaşılırlığı sebebiyle en çok kullanılan ölçek, yine Saaty (1980) tarafından önerilmiş olan ve "Temel Ölçek" adıyla da bilinen 1-9 lineer ölçeğidir.

İkili karşılaştırmalar sonucu elde edilen sayısal değerler, "İkili Karşılaştırma Matrisi" (İKM) adı verilen kare matrislerde tutulur. İKM'lerde tutulan bu sayısal değerler, karşılaştırılan bütün elemanların kendi grupları içindeki yerel önemlerini (ağırlıklarını) hesaplamak için kullanılır. Yerel ağırlıkları hesaplamak için kullanılan en yaygın yöntemler Özdeğer (Eigenvalue) Yöntemi ve Logaritmik En Küçük Kareler (ya da başka bir adıyla Satır Geometrik Ortalama) Yöntemi'dir.

Yerel ağırlıklar belirlendikten sonra karar vericinin değerlendirmelerinin tutarlılığı bütün (ya da en azından tutarlılığı şüpheli olan) İKM'ler için değerlendirilir. En yaygın bilinen değerlendirme yöntemi Özdeğer Yöntemi; ikinci en bilinen yöntem ise Geometrik Tutarlılık Endeksi'dir. Doğası gereği AHS belirli derecede tutarsızlık barındırır. Fakat bu tutarsızlık belirli sınırlar içinde kaldığı sürece kabul edilebilir olarak değerlendirilir. Eğer bir İKM'nin tutarsız olduğu (tutarsızlık değerinin sınırlar dışında olması) tespit edilirse, karar vericinin İKM'yi –tutarsızlık Kabul edilebilir seviyeye inene ya da karar verici daha fazla değişikliğin kendi fikirlerini yansıtmayacağını belirtene kadar- tekrar değerlendirmesi istenir. Yakın tarihli bir çalışmada (Mazurek ve Perzina, 2017) sunulan, insanların az sayıda elemanı değerlendirirken bile büyük oranda tutarsız gösterdiği sonucu göz önünde bulundurulduğunda, tekrar değerlendirme işleminin ciddi oranda zaman alıcı bir işlem haline geldiği görülmektedir.

Tutarlılık değerlendirmesinden sonra en son adım olan ağırlıkların bütünleştirilmesi işlemi yapılır. Bu adımda en alt seviyede bulunan her bir ölçütün ana hedefe ulaşmak için ne kadar önemli olduğu tespit edilir. Bütünleştirme işlemi sıklıkla Toplamsal Bütünleştirme Yöntemi ile yapıldığı gibi Çarpımsal Bütünleştirme Yöntemi de kullanılmaktadır.

Bir İKM'deki tutarlılık genel anlamda AHS'nin "geçişlilik" önermesine bağlanabilir. Geçişlilik kavramı ikiye ayrılır:

- Nitel Geçişlilik: Eğer A B'ye tercih ediliyor ve B de C'ye tercih ediliyorsa A'nın C'ye tercih edilmesi gerekir.
- Nicel Geçişlilik: Eğer A B'ye göre 3 kat tercih ediliyor ve B de C'ye göre 2 kat tercih ediliyorsa A'nın C'ye göre 6 kat tercih edilmesi gerekir.
- Geçişlilik önermesine göre bir İKM'nin tam olarak tutarlı olabilmesi için bu İKM'nin nicel geçişlilik kuralına uyması gerekir. Diğer yandan, eğer bu İKM nicel olarak geçişliyse, aynı zamanda nitel olarak da geçişlidir.

İKM'lerin tutarsız olması istenmeyen bir durum olduğu gibi karmaşık bir karar verme sorununda çok sayıda tutarsızlığın düzeltilmeye çalışılması külfetli bir iştir. Tutarsızlık sorununa sayısal ikili karşılaştırma ölçeklerini göz önünde bulundurarak yaklaşıldığında iki tip tutarsızlık görülmektedir. Birinci tip tutarsızlık, ölçeğin bir üst ve alt limite sahip olmasından kaynaklanmaktadır. Örneğin; i-j elemanları arasındaki sayısal karşılaştırma değeri "$a_{ij} = 4$" ve j-k elemanları arasındaki sayısal karşılaştırma değeri "$a_{jk} = 5$" olduğunda, i-k elemanları arasındaki sayısal karşılaştırma değerinin nicel geçişlilik önermesine göre "$a_{ik} = 20$" olması beklenir. Bu durumda, üst sınırı 20'den küçük olan bir ölçeğin tutarsızlık göstermesi kaçınılmazdır. Araştırmamızda, AHS için önerilmiş olan sayısal ölçeklerin çoğunluğunun bu durumdan ciddi derecede olumsuz etkilendiği tespit edilmiştir. Fakat, üst limiti olan herhangi bir sayısal ölçeğin bu tip bir tutarsızlıktan zorunlu olarak etkileneceği de belirtilmiştir.

İlkinden daha etkili olduğunu düşündüğümüz bir diğer sayısal ölçek kaynaklı tutarsızlık sebebi ise ölçek tarafından tanımlanmış olan sayısal değerlerin çarpımsal kombinasyonlarının ölçek sınırları içinde kalmasına rağmen, ölçek tarafından tanımlanmamış bir sayısal değere eşit olmasıdır. Başka bir deyişle, "$a_{ik} = a_{ij} * a_{jk}$" eşitliğinin sayısal değeri ölçek sınırları içinde olsa bile "a_{ik}" değeri kullanılan ölçek tarafından tanımlanmamış olabilir. Örneğin, $a_{ij} = 1/3$ ve $a_{jk} = 4$ durumunda a_{ik} değerinin $4/3$ olması beklenmektedir. AHS için önerilmiş olan sayısal ölçeklerin (Geometrik hariç) bu tip bir tutarsızlıktan büyük oranda etkilendikleri tespit edilmiştir.

Araştırmamızda, sayısal ölçek kaynaklı tutarsızlıkları azaltacağı düşünülen ve "Üstel Ölçek" olarak adlandırılan yeni bir ölçek önerisi getirilmiştir. Üstel ölçek kavramı ilk olarak Lootsma (1989) tarafından ortaya atılmış olup, ilgili araştırmada sunulan genel formüldeki katsayılar bir temele bağlı olarak belirlenmemiş ve bu nedenle önerilen formül bir sayısal ölçeğe dönüştürülmemiştir. Bu araştırmada ise katsayıların belirlenmesi aşamasında yaygın olarak bilinen "Fibonacci Dizisi" kullanılmış ve bu diziyi temel alan bir sayısal ölçek önerilmiştir. Beklendiği gibi, Üstel Ölçek'in ölçek kaynaklı tutarsızlıklar konusunda yüksek bir performans sergilediği tespit edilmiştir.

Daha sonra, detaylı simülasyonlarla, daha önce önerilmiş olan ölçekler ve araştırmamızda önerilen Üstel Ölçek, Saaty tarafından önerilen Temel Ölçek ile karşılaştırılmıştır. Karşılaştırılan ölçekler belirli performans ölçütlerine göre değerlendirilmiş ve özellikle tutarlılık ile ilgili ölçütler göz önünde bulundurulduğunda Kare Ölçek'in diğerlerinden daha üstün olduğu tespit edilmiştir. Üstel ve Geometrik ölçekler Kare Ölçek'i yakından takip ederken Ters Lineer ve Dengeli ölçeklerin ortalama bir performans gösterdiği gözlemlenmiştir. Logaritmik ve Karekök ölçeklerin ise, Temel Ölçek'le kıyaslandığında, tutarlılık ile ilgili ölçütlerde diğer ölçeklere göre ciddi derecede düşük performans gösterdikleri tespit edilmiştir. Elde edilen sonuçlarda dikkat çeken bir nokta ise yüksek performans gösteren Üstel, Geometrik ve Kare ölçeklerinin hepsinin üst sınırlarının Temel Ölçek'te kullanılan 9'dan daha büyük olmasıdır. Benzer şekilde orta derecede performans gösteren Ters Lineer ve Dengeli ölçeklerin üst sınırlarının 9'a eşit; düşük performansa sahip Logaritmik ve Karekök ölçeklerin üst sınırlarının ise 9'dan küçük olduğu gözlemlenmiştir. Bu bağlamda, bir ölçeğin alt ve üst sunurlarının, o ölçekle oluşturulam İKM'lerin tutarlılığı üzerinde ciddi derecede etkili olduğu ve genel anlamda sınırı 9'dan büyük olan ölçeklerin –aynı sözel değerlendirme seti baz alındığında– daha yüksek performans gösterdiği sonucuna varılmıştır.

Araştırmanın son bölümünde ise bütün ölçekler aynı örnek karar verme uygulaması üzerinde iki farklı senaryoda (Temel Ölçek'e göre tutarsız ve tutarlı İKM senaryoları) değerlendirilmiştir. Tek bir örnek üzerinden çok genel bir çıkarım yapılmasının sağlıklı olmadığına inanmakla birlikte, üst sınırları 9'dan büyük olan ölçeklerin (Üstel, Geometrik ve Kare) beklendiği şekilde en önemli alternatif ağırlıklarını diğer ölçeklere göre daha yüksek belirledikleri görülmüştür. Ek olarak Üstel ve Geometrik ölçeklerle oluşturulan İKM'lerin her iki senaryoda da tutarlı olmasına karşın Temel ve Kare ölçeklerle oluşturuan İKM'lerin sadece ikinci senaryoda tutarlı oldukları gözlemlenmiştir.

Geometrik Ölçek ile Üstel Ölçeğin değerlendirme ölçütlerimize göre performansları birbirine yakın olmakla birlikte, Ji ve Jiang (2003) tarafından belirtildiği gibi üst sınırı 256 olan bir ölçeğin kullanımının AHS'nin homojenlik önermesini ciddi şekilde ihlal ettiğini düşünüyoruz. Bu bağlamda, Geometrik Ölçek yerine Üstel Ölçek'in kullanılmasının daha uygun olduğuna inanıyoruz.

Kare Ölçek Harker ve Vargas (1987) tarafından önerilmiş olup yine aynı araştırmada tek bir örnek üzerinden eleştirilmiş ve Temel Ölçek'in daha uygun olduğu çıkarımı yapılmıştır. Bu araştırmada uygulanan örnek, Saaty (1980) tarafından yapılan araştırmada Philadelphia ile altı farklı şehir (Kahire, Tokyo, Şikago, San Fransisko, Londra ve Montreal) arasındaki normalize gerçek uzaklıkları tahmin etmek için kullanılmıştır. Harker ve Vargas (1987) tarafından yapılan araştırmada karşılaştırılan bütün sayısal ölçekler için aynı sözel ölçek kullanılmış ve bunun sonucunda oluşturulan İKM'ler ile belirlenen ağırlıklar, Philadelphia ile bahse konu şehirler arasındaki normalize gerçek mesafelerle karşılaştırılmıştır. Araştırmanın sonucu olarak Temel Ölçek'le oluşturulan ağırlıkların normalize gerçek mesafelerle yüksek korelasyona sahip olduğu ve bu nedenle Temel Ölçek'in diğerlerinden daha iyi olduğu öne sürülmüştür. Fakat araştırmacılar bu çalışmada önemli bir noktayı atlamış; hangi sözel değerlendirmenin hangi sayısal puana karşılık geldiğini bilen mantıklı bir karar vericinin hem Temel hem de Kare ölçekler için aynı sözel değerlendirme ölçeğini kullanacağı varsayımında bulunmuşlardır. Bizim öngörümüze göre ise, sayısal ölçekleri bilen ve karşılaştırdığı elemanlar hakkında fikir sahibi olan mantıklı bir karar vericinin, sayısal ölçek değiştiğinde kendi değerlendirmelerini yeni ölçeğe göre ayarlayacağı yönündedir. Dolayısıyla bütün sayısal ölçekler için aynı sözel değerlendirmeler kullanılarak değerlendirme yapılmasına dikkatli bir şekilde yaklaşılması gerektiğine inanıyoruz. Bu bağlamda Harker ve Vargas (1987) tarafından önerilmiş ve eleştirilmiş olan Kare Ölçek'in belirtildiği kadar elverişsiz olmayabileceğine; hatta daha tutarlı İKM'lerin oluşturulması sayesinde karar vericilerin İKM'leri tekrar tekrar değerlendirmesinin önüne geçilerek süreç için daha çok fayda sağlanabileceğine inanıyoruz.

Bu sınırlama aynı zamanda bizim araştırmamızda da önümüze çıkıyor. Yapılan simülasyonlarda bir İKM öncelikle Temel Ölçek'e göre rasgele belirlenip daha sonra bu İKM'deki sayılar karşılaştırılan ölçekteki karşılıklarıyla ikinci bir İKM olarak oluşturulmuştur. Başka bir deyişle, simülasyonlarda Temel Ölçek ve karşılaştırılan ölçekle oluşturulan İKM'ler için aynı sözel değerlendirmeler kullanılmıştır. Bu bağlamda, farklı sayısal ölçekler kullanıldığı durumlarda karar vericilerin sözel ölçekleri nasıl algıladığı konusunda bir araştırma yapılabilir. Yani, "Bir karar verici, sayısal ölçek değiştiği zaman sözel değerlendirmelerini değiştirir mi?" sorusuna deneysel olarak cevap aranabilir.

Araştırmamızın bir başka kısıtı ise ağırlıkların belirlenmesinde sadece Özdeğer Yöntemi'nin kullanılmasıdır. Öncül simülasyonlarımızda tespit ettiğimiz üzere, çalışma içinde açıklanan Yaklaşık Değer Yöntemi kullanılarak belirlenen ağırlıklar

Özdeğer Yöntemi'yle belirlenen ağırlıklarla yüksek oranda benzerlik göstermiştir. Bu nedenle belirtilen iki yöntemden sadece Özdeğer Yöntemi kullanılmıştır. Ek bir araştırma olarak ikinci en çok kullanılan ağırlık belirleme yöntemi olan Logaritmik En Küçük Kareler (Satır Geometrik Ortalama) Yöntemi kullanılabilir.

Bir başka kısıt ise tutarlılık değerlendirmesinde, bilgisayımsal kapasite ve hız göz önünde bulundurularak, yalnızca Özdeğer Yöntemi'nin kullanılmış olmasıdır. Buna ek olarak yapılabilecek yeni bir araştırmada tutarlılık değerlendirmesi için Geometrik Tutarlılık Endeksi kullanılabilir. Aynı zamanda ağırlık belirleme yöntemleri (Özdeğer ve Logaritmik En Küçük Kareler) ile tutarlılık değerlendirme yöntemlerinin (Özdeğer ve Geometrik Tutarlılık Endeksi) bu araştırmada uygulanmamış olan eşleştirmeleri uygulanabilir.

Simülasyonlarımızda tamamen rasgele İKM'ler oluşturmuş olmamıza rağmen yalnızca tutarlılık değerleri belirli sınırların altında kalan İKM'ler değerlendirmeye alınmıştır. Daha geniş bir tutarlılık değeri aralığını kapsayacak şekilde yapılabilecek bir araştırma, kullanılan ölçeklerin İKM'lerin tutarlılığı üzerinde etkisi konusunda daha detaylı bilgi verebileceğini düşünüyoruz.

Geçmişte yapılan çalışmalarda tek örnek üzerinden ölçek değerlendirmesi yerine geniş çaplı simülasyonlarla sayısal ikili karşılaştırma ölçeklerinin genel özelliklerini ortaya çıkarma amacı güdülmesinin, bizim araştırmamızı geçmişte yapılan araştırmalardan farklı kıldığına inanıyoruz. Ek olarak, Ji ve Jiang (2003) tarafından yapılmış olan ve ölçek kaynaklı tutarsızlık özelliğini ipuçlarının, muhtemel bütün kombinasyonları içeren çalışmamızda daha geniş kapsamlı ve yol gösterici şekilde incelendiğini düşünüyoruz.

Son olarak, AHS yöntemi kullanılarak yapılan tekil uygulamalar üzerinden genel sonuçlara varılmasının uygun olmadığı görüşümüzü belirtmek istiyoruz. Simülasyonlarımızın da gösterdiği üzere Üstel, Geometrik ve Kare ölçeklerin Temel Ölçek'ten daha tutarlı sonuçlar verdiği durumlar çoğunlukta olmakla birlikte, geri kalan yadsınamayacak sayıdaki durumlarda ise Temel Ölçek'in daha tutarlı sonuçlar verdiği gözlemlenmiştir. Bu açıdan bakıldığında, tekil bir uygulamanın okuyucuyu

Üstel Ölçek'in daha iyi olduğu bir duruma yönlendirebileceği gibi, Temel Ölçek'in daha iyi sonuçlar verdiği bir duruma da yönlendirebilir. Bu bağlamda, yüksek sayıda durum içeren simülasyonlar üzerinden yapılacak değerlendirmelerin daha güvenilir sonuçlar vereceğine inanıyoruz.

D. TEZ İZİN FORMU / THESIS PERMISSION FORM

ENSTİTÜ / INSTITUTE

- [ ] **Fen Bilimleri Enstitüsü** / Graduate School of Natural and Applied Sciences
- [x] **Sosyal Bilimler Enstitüsü** / Graduate School of Social Sciences
- [ ] **Uygulamalı Matematik Enstitüsü** / Graduate School of Applied Mathematics
- [ ] **Enformatik Enstitüsü** / Graduate School of Informatics
- [ ] **Deniz Bilimleri Enstitüsü** / Graduate School of Marine Sciences

YAZARIN / AUTHOR

Soyadı / Surname : Yıldırım
Adı / Name : Boğaç Can
Bölümü / Department : İşletme/Business Administration

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English) : A NEW PAIRWISE COMPARISON SCALE FOR ANALYTIC HIERARCHY PROCESS

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans** / Master [x] **Doktora** / PhD [ ]

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. [ ]

2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two year**. * [ ]

3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. * [x]

***** *Enstitü Yönetim Kurulu Kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir.*
A copy of the Decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.

Yazarın imzası / Signature .... **Tarih** / Date 18.01.2018