MODELING THE SUPPLIER SELECTION PROCESS FOR MECHANICAL PARTS

by

FATİH HABACI

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfilment of
the requirements for the degree of Master of Science.

Sabancı University

July 2023

ABSTRACT

MODELING THE SUPPLIER SELECTION PROCESS FOR MECHANICAL PARTS

FATİH HABACI

INDUSTRIAL ENGINEERING M.S. THESIS, JULY 2023

Thesis Supervisor: Assoc. Prof. Dr. KEMAL KILIÇ

Keywords: Supplier Selection, Multi-Criteria Decision-Making, Analytic Hierarchy Process, Manufacturing, Distributed Decision-Making with Missing Data

As the complexity of outsourcing operations increases, the organizations' procurement decisions get more complicated. Although there are various previous works about supplier selection for different industries, there exists no widely accepted universal model. This is due to the fact that organizations' decisions are shaped differently according to their objectives and business environments. Moreover, even the same organization may need different selection criteria and methods for different products. In this study, a supplier selection model is developed for an organization that needs diverse mechanical parts which are produced by machining processes. We consider an organization that outsources the production operations of the mechanical part which are designed by the organization. Since the complexity of the operations is high and the parts are almost unique, manufacturability is a critical issue. Therefore, besides widely used criteria, the technical competencies of the suppliers are considered as well. The scope of this study is to determine the shortlist of the suppliers to be asked for a quote, rather than evaluation of the quotes. The motivation here is that although having more quotes is beneficial for the organization, it is a costly operation. Therefore, determining the ideal set of suppliers for a product is valuable. In this study, an ideal criteria set is determined upon interviews with professionals. Then, importance levels (i.e., weights) of the criteria are determined by the procurement staff. As the next step, evaluation procedures are defined for criteria. For the evaluation, extensions such as distributed decision-making and decision-making with missing data are used. Afterward, a fitting score is found for supplier and mechanical part pairs, on which supplier recommendations are made. For fitting scores, besides the cost and benefit perspectives on the value functions, the target value approach is utilized.

ÖZET

MEKANİK PARÇALAR İÇİN TEDARİKÇİ SEÇİMİ MODELLEMESİ

FATİH HABACI

ENDÜSTRİ MÜHENDİSLİĞİ YÜKSEK LİSANS TEZİ, TEMMUZ 2023

TEZ DANIŞMANI: Doç. Dr. KEMAL KILIÇ

Keywords: Tedarikçi Seçimi, Çok Ölçütlü Karar Verme, Analitik Hiyerarşi Süreci, Üretim, Eksik Veri ile Dağıtık Karar Verme

Organizasyonların dış kaynak kullanım operasyonlarının karmaşıklığı arttıkça, satın alma kararları da daha karmaşık hale gelmektedir. Literatürde Tedarikçi Seçimi konusunda birçok geçmiş çalışma olmasına rağmen, yaygın şekilde kabul edilip her organizasyona uygun olan bir model bulunmamaktadır. Bunun nedeni, organizasyonların satın alma kararlarının kendi amaçları ve iş dünyalarının gerekliliklerine göre değişkenlik göstermesidir. Bunun yanında, tek bir organizasyon bile farklı ürünler için farklı seçim kriterlerine ve metotlarına ihtiyaç duyabilmektedir. Bu çalışmada, talaşlı imalat ile üretilen oldukça çeşitli mekanik parçalara ihtiyaç duyan bir organizasyon için bir tedarikçi seçimi modeli geliştirilmiştir. Çalışma konusu olarak seçilen organizasyon, kendi tasarladığı mekanik parçaların üretimi için dış kaynak kullanmaktadır. Operasyonların karmaşıklığının yüksek olması ve parçaların neredeyse benzersiz olması itibariyle, üretilebilirlik bu çalışma için kritik bir husus olacaktır. Bu bağlamda, yaygın olarak kullanılan kriterlerin yanında, tedarikçilerin teknik yetkinlikleri de ele alınacaktır. Bu çalışmanın kapsamı teklifleri değerlendirerek en iyi alternatifi seçmek değil, teklif istenecek firmaların kısa listesini elde edebilmektir. Buradaki motivasyon, fazla teklif almanın firmaya getirdiği faydaların yanında yol açtığı operasyonel maliyet ve iş yüküdür. Bu yüzden teklif istenecek firma

kısa listesinin doğru belirlenmesi kıymetlidir. Bu çalışmada, ilgili alanda çalışan kişilerle yapılan mülakatlar neticesinde ideal bir kriter kümesi sunulmaktadır. Sonrasında bu kriterlerin önem seviyeleri de (ağırlıkları) yine satın alma personelleri tarafından belirlenmektedir. Sonraki adımda ilgili kriterler için tedarikçilerin değerlendirme yöntemleri tanımlanmaktadır. Bu değerlendirme yöntemlerinde dağıtık karar verme ve eksik veriyle karar verme gibi uzantılar kullanılmaktadır. Sonrasında, tedarikçiler ile mekanik parçalar arasında bir uyum skoru belirlenerek tedarikçi önerileri sunulmaktadır. Uyum skorları için fayda ve maliyet yaklaşımlarının yanında, hedef değer yaklaşımı da kullanılmaktadır.

ACKNOWLEDGEMENTS

I would like to thank Dr. Kemal KILIÇ and Dr. Emre ÖZLÜ for their support, inspiration, and all the learning opportunities. I would also like to thank my committee members, Dr. Erhan Budak and Dr. Faran Ahmed.

I would also like to thank the Scientific and Technological Research Council of Turkey (TÜBİTAK) for their support in this study.

To the people who are always by my side

My parents, my brother, and my beloved wife

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AHP Analytic Hierarchy Process

AI Artificial Intelligence

AIJ Aggregating Individual Judgments

AIP Aggregating Individual Priorities

ANP Analytic Network Process

BWM Best-Worst Method

C.I. Consistency Index

C.R. Consistency Ratio

CAM Computer Aided Manufacturing

CMM Coordinate Measuring Machine

CNC Computerized Numerical Control

DEA Data Envelopment Analysis

DEMATEL Decision Making Trial and Evaluation Laboratory

LP Linear Programming

MACBETH Measuring Attractiveness by a CategoricalBased Evaluation Technique

MAUT Multi-Attribute Utility Theory

MCDM Multi-Criteria Decision-Making

MNL Multinomial Logit Model

MP Mathematical Programming

PROMETHEE Preference Ranking Organization Method for Enrichment Evaluation

R.I. Random Index

SEM Sparse Eigenvector Method

SGCM Sparse Group Comparison Matrix

TCO Total Cost of Ownership

TOPSIS The Technique for Order of Preference by Similarity to Ideal Solution

TRIR Total Recordable Incident Rate

VIKOR VlseKriterijumska Optimizacija I Kompromisno Resenje (in English: Multi-Criteria Optimization and Compromise Solution)

Chapter 1

INTRODUCTION

As the complexity of outsourcing operations gets increased, the procurement decisions of the companies get more complicated. This complexity arises from multiple factors being included in the decision-making process of choosing a suitable supplier. In this sense, supplier selection appears as an application field of the Multi-Criteria Decision-Making (MCDM) discipline. Although there exists a wide variety of work about supplier selection for different industries, there is still room for further studies such as new methods and hybrid usage of the current methods due to the immense diversity in the problem setting.

In this thesis, a supplier selection problem is studied within a make-to-order environment, and the company whose supplier selection process is the subject of this study will be referred to as the organization in the rest of the thesis. The organization deals with a case in which purchase requisitions are placed for mechanical parts which are designed within the company. Since these items are not commercial on-the-shelf and design processes are held within the organization, the required items are almost unique. Therefore, the manufacturability of the mechanical part by suppliers is one of the challenges in this study. Hence, the proposed method should also take the specifications and complexity of the mechanical parts into consideration. Consequently, the first research question is how to enable the supplier selection method to be sensitive to the features of the mechanical parts. In a more general expression, how the value functions should be defined so as to comply with the decisions of procurement staff is the first research question.

Although the supplier selection studies in the literature frequently focus on the evaluation process of the quotes, there is a preceding stage of supplier selection, which is selecting the suppliers which are requested for a quotation. In other words, there is a larger set of suppliers which are possible candidates for the order, and some of them are selected as the shortlist of the suppliers which are possibly best fit for the outsourcing of the requested mechanical parts. One crucial difference here is that the decision has to be made under missing information which is defined with the quotation stage. Hence, the second research question is how to make a wise decision with the lack of some critical information such as lead time and price.

As it is stated in the previous paragraph, the first stage of the supplier selection decision is made among a large set of alternatives. Unfortunately, procurement staff as decision-makers do not have an unlimited capacity or resources to process the required information since they are human beings. For instance, the organization might lose a price advantage if a procurement staff does not think of one of the appropriate suppliers as a candidate for a mechanical part. Moreover, the knowledge and experience of the organization might be lost when there is a high turnover rate, so new staff may not know the suppliers well. Therefore, this decision-making process should be enriched with a data-driven decision-support system to utilize more information while making the decision.

In the literature, a general flow of a supplier selection process is as follows. There are multiple suppliers which send a quote for a product. The decision-maker defines related criteria and determines their weights. Then, the performances of the suppliers are evaluated with respect to the criteria. Afterward, the supplier with the highest point is selected. Nonetheless, the underlying unrealistic assumption here is that the decision-maker is willing to make all the comparisons for criterion pairs every time. When the number of orders is large, it may not be even feasible. Therefore, a comprehensive methodology should be developed for a set of products rather than a single product in such a system.

Including but not limited to the points discussed above, one of the most important steps of this decision-making process is determining the criteria. While determining the criteria, the measurability and evaluability of them should also be assessed. Therefore, rather than general and most commonly used criteria such as quality and delivery, the criteria being used have to address a numerical performance value for applicability. In other words, the metric for quality and delivery performance should be stated clearly, for instance. There are several motivations for preferring criteria that are either measurable or evaluable. The first one is that some methods for evaluation of the alternatives might not be feasible such as pairwise comparison due to the largeness of the candidate supplier set. Hence, the performance data of the suppliers should be converted into a normalized scale for the sake of simplicity of use. Another motivation is to enable the decision maker to process a big set of information considering each supplier is associated with multiple purchase orders in which different performance levels are observed. Last, adopting a common evaluation approach for every decision maker is desired for improving the consistency of decisions in the system since there are multiple procurement staff members. These motivations can be transformed into a systematic by either defining a rule to evaluate the performance of the alternatives or reaching a consensus on the performance of the suppliers by distributed decision-making. This discussion leads to two different research questions of which one is what the ideal criteria set is. The fourth research question is how to combine distributed decisions into a final decision.

In the evaluation process, ordinal and numeric data can be converted into a numerical scale, and they can be used in the value functions of the decision maker easily. Nonetheless, there are some criteria that cannot be associated with a numerical scale. In this case, the decision maker must make pairwise comparisons for each alternative. However, in such an environment where the number of alternatives is high, having all the pairwise comparisons is impossible due to the combinatorially growing number of questions. Thus, the last research question appears as what is an ideal way to evaluate the alternatives with distributed decisions under the sparsity of data.

This thesis is organized as follows. In Section 2, findings on past studies regarding the supplier selection field are given. In Section 3, the environment of the problem is described. In Section 4, the proposed solution is presented. In Section 5, the results of the studies are discussed, and further research opportunities are stated.

Chapter 2

LITERATURE REVIEW

Supplier selection is an attractive study field in that a big diversity in various manners is observed. For example, researchers from diverse disciplines put their effort from different points of view. As another demonstrated diversity, various criteria, and methods are proposed for different problem settings from the same topic. Moreover, the focus of these studies also varies as some are focusing on the evaluation of the tenders while others studying the decision of adding a new supplier to the network. Due to the countless publications in this field arising from this diversity, statements on the whole of this literature would be overconfident since it is not possible to cover all this extensive literature. However, this literature review starts with a shallow statistic about this literature, which is the fact that approximately 139,000 results are returned by Google Scholar when the exact phrase ”Supplier Selection” is searched by advanced search, by the end of June 2023.

Beil (2011) lists three reasons why new suppliers are important. The first reason is that a supplier might be using a novel production technology which decreases the costs. Secondly, a supplier might have a more advantageous cost structure than others. The third reason is the several motivations of the buyer such as building a more robust supply chain by reducing supply risks, having a price advantage by increasing competitiveness, and other strategic objectives. In this manner, one takeaway from this publication is to add new suppliers to the network for benefiting in diverse ways. However, new suppliers bring more risks to the firms which make their decisions only on historical experiences since an observation means an order placed to the new supplier. Therefore, supplier selection studies to build a system gains more meaning because these studies enable the buyer to have objective judgments on the new supplier.

MCDM methods can be utilized in order to rank, score, and outrank the alternatives over often conflicting criteria. In this sense, MCDM methods are widely used in supplier selection problems as there are multiple criteria in this decision-making process such as lead time, price, and quality for which generally one dominant alternative does not exist.

Outsourcing non-core operations can be benefited by focusing on the firm's core business and utilizing the competitive advantage of the supplier (Kenyon et al., 2016). The important part here is making a suitable supplier selection. Nonetheless, this decision-making process highly depends on the problem setting, hence there is no standard way to make a suitable supplier selection (Taherdoost and Brard, 2019).

For benefiting from the existing literature and designing the solution accordingly, the results of the studies on the criteria and methods will be presented in the next subsections. In addition, there will be a detailed literature review of the Analytic Hierarchy Process (AHP) which occupies an important place in this study.

2.1 Criteria for Supplier Selection

In Sarkis and Talluri (2002), 31 sub-factors are incorporated for supplier selection under seven categories of which four of them are strategic performance metrics and three are organizational factors. Four categories of strategic performance metrics are cost, quality, time, and flexibility. Besides, culture, technology, and relationships are three organizational factors. One key finding in this paper is that the supplier selection process is sensitive to product type.

Huang and Keskar (2007) present 101 performance metrics for supplier evaluation in 7 categories, and these metrics are also categorized with respect to product type (i.e. make to stock, make to order, and engineer to order) and location of the supplier. The seven main categories in this paper are reliability, responsiveness, flexibility, cost and financial, assets and infrastructure, safety, and environment.

Verma and Pullman (1998) conducted an empirical study with 58 managers. The purpose of this study is to interpret the priorities of 5 attributes of the suppliers, which are cost, quality, lead-time, on-time delivery, and flexibility. In the first part, a five-point Likert-type scale is provided to the participants, and they are asked to evaluate the importance of the attributes. In the second part, 16 different supplier profiles are generated with respect to the attributes, and the participants' preferences are obtained. Based on these preferences, a multinomial logit model (MNL) is used to explain the decisions with respect to the performance levels of the suppliers over the criteria set. According to the significance level of 0.05, all criteria but flexibility appeared to be significant in affecting the decisions. An interesting result here is that the Cost criterion has the highest absolute value of the coefficient in the logit model, whereas it is ranked in the 3rd place in the evaluation by the Likert scale. Another important finding is that On-time delivery appears to be more important than Lead-time in both parts. In other words, having long lead times is more acceptable than being not loyal to the committed date. The results of this study are outlined in Table 2.1.

Table 2.1: Results of the Empirical Study (Verma and Pullman, 1998)

	Likert Scale Scores		**Parameter Estimates in MNL**	
Criterion	**Mean**	**Median**	**Coefficient (beta)**	**p-value**
Cost	3.96	4	-0.577	0.000
Quality	4.56	5	0.384	0.001
Lead-time	3.87	4	-0.287	0.013
On-time delivery	4.14	4	0.416	0.001
Flexibility	3.22	3	0.045	0.695

Barbarosoglu and Yazgac (1997) defined 3 categories of primary objectives in their supplier selection study, which are Performance Assessment, Quality System Assessment, and Business Structure/Manufacturing Capability Assessment in the order of relative importance. The priorities of these primary objectives are found with AHP as 0.625, 0.136, and 0.238, respectively. There are 16 criteria just below these categories in this five-level hierarchy of criteria. According to the results of the AHP study, the criteria among these 16 which have more importance than average are Shipment Quality, Delivery, Cost Analysis, and Management Commitment. Their weights and associated primary objectives are given in Table 2.2.

Table 2.2: Criteria with a Score Higher Than Average (Barbarosoglu and Yazgac, 1997)

No.	**Primary Objective**	**Criterion**	**Priority**
1	Performance Assessment	Shipment Quality	0.268
2		Delivery	0.268
3		Cost Analysis	0.089
4	Quality System Assessment	Management Commitment	0.094
		Average	**0.0625**

Kannan and Tan (2002) defined two different criteria sets for supplier selection and assessment. In the supplier selection part, 30 criteria are included in the list which is mailed to the participants. In this mail, the participants are asked for evaluating these criteria with respect to a five-point scale. In this study which 411 surveys are used, Ability to meet delivery due dates and Commitment to quality criteria appeared to be the two most important criteria with 0.02 points of difference between them. Among 30 criteria, 8 of them with a higher score than 4.00 are provided in Table 2.3.

Table 2.3: Criteria with a Score Higher Than 4.00 (Kannan and Tan, 2002)

No.	Selection Criteria	Mean Score
1	Ability to meet delivery due dates	4.62
2	Commitment to quality	4.60
3	Technical expertise	4.25
4	Price of materials, parts, and services	4.16
5	Honest and frequent communications	4.11
6	Reserve capacity or the ability to respond to unexpected demand	4.08
7	Industry knowledge	4.06
8	Financial stability and staying power	4.03
	Average of Mean Scores of 30 Criteria	**3.65**

Chan and Chan (2004) developed a supplier selection model for the advanced technology industry. In order to select the supplier, 6 main criteria are defined in this study, which are cost, delivery, flexibility, innovation quality, and service. There are 20 subcriteria just below these six in the hierarchy of criteria. In this study, an AHP study is conducted with 26 participants to derive the priorities of the main criteria and subcriteria. As a result, quality appears to be the most important criterion followed by delivery and cost. These 3 criteria contribute to the objective with a total weight of more than 0.75, and the remaining three have a total weight of less than 0.25. The subcriteria with a weight of more than average, which is 0.05, and their associated main criteria are presented in Table 2.4.

One criterion set gaining more and more importance is the environmental criteria. Rezaei et al. (2016) combined traditional criteria such as price and environmental criteria such as sustainability performance. While the four most important criteria are traditional criteria such as price, delivery cost, delivery time, and quality, which make up 76 percent of the total weight, it is important to note that the environmental success of suppliers is reflected in the supplier selection process.

Table 2.4: Criteria with a Score Higher Than Average (Chan and Chan, 2004)

No.	Criterion	Global Weight of Criterion	Subcriterion	Global Weight of Subcriterion
1	Quality	0.403	Conformance to Specifications	0.164
2			Product Reliability	0.139
3			Product Durability	0.101
4	Delivery	0.186	Delivery Reliability	0.065
5	Cost	0.167	Competitive Pricing	0.068
6			Total Cost	0.059
			Average	**0.05**

Several conclusions for the literature review of supplier selection criteria can be stated. One of them is that supplier selection is a complex problem since there are many criteria at the lowest level of the hierarchy. However, the decision-makers may not provide precise judgments after a point, so an attempt to add more resolution to the problem might result in over-complexity. Another conclusion is that the importance levels of the criteria highly depend on the case. Although there are similarities in the criteria sets among papers, there is no consensus on their priorities of them except for a few widely accepted criteria such as delivery and quality. Therefore, the initial criteria set can be derived from the literature, but finalizing must be tailor-made to the organization. Moreover, variations in procurement decisions may appear for different product groups even in the same organization. Hence, a requirement of such sensitivity should be investigated. Last, some new criteria emerge while some of them are losing attention in order to comply with the necessities of the dynamic environment of the operations. Hence, studies in this field will not be unrequited since these sets are subject to change even if there were some ideal criteria set which is applicable to every case for today's circumstances.

2.2 Methods for Supplier Selection

Nydick and Hill (1992) utilized AHP to select the best supplier. For this objective, every criterion pair is compared to derive the relative importance of the criteria. Then, every supplier pair is compared with respect to every criterion. In this way, the total value of the supplier is obtained by combining the importance of the criteria and the performance of the suppliers with respect to the associated criterion. By this method, a total weight of 1 is divided into the alternatives, and the best alternative can be found as a reflection of the comparisons of the decision maker.

AHP is not only used on its own but also commonly used in hybrid methods in decision-making processes. Not limited to MCDM methods, AHP is also used with other decision-making techniques such as Mathematical Programming (MP) as well. For instance, Ghodsypour and O'Brien (1998) used AHP to derive the weights of the criteria and then utilized LP (Linear Programming) in order to allocate the total order amount among the suppliers for maximizing the total value while complying with the constraints such as capacity and quality.

Another set of decision-making methods, used with AHP in hybrid solutions is the examples of compromise methods. Fazlollahtabar et al. (2011) derived the priorities of the criteria with AHP and evaluated the performances of the alternatives with The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) in the supplier selection context. Afterward, order quantities are allocated to the suppliers with an MP technique. As another similar approach, Prasad et al. (2016) derived the priorities of the criteria with AHP, then selected the best supplier by using VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR, in English: Multi-Criteria Optimization and Compromise Solution). As an observation, AHP is commonly used for derivation of the weights of the criteria even if alternatives are not evaluated with AHP.

Gencer and Gürpinar (2007) used Analytic Network Process (ANP) to deal with a problem in which criteria and alternatives have influences among them. In this study, the independence assumption and need for building a hierarchy is avoided thanks to the usage of ANP. The problem studied in this publication is a supplier selection case for an electronic firm.

With a similar subject, Weber (1996) applied Data Envelopment Analysis (DEA) to evaluate the performances of the suppliers. DEA is an MP technique that evaluates the efficiency of the alternatives based on the inputs and outputs regarding these alternatives. Liu et al. (2000) used a simplified DEA model to detect inefficient suppliers and reduce the number of suppliers by transferring the orders allocated to inefficient ones to others.

Ellram (1995) examined the Total Cost of Ownership (TCO) method for procurement decisions. With its complex nature, the TCO method provides a wider understanding of the total cost of a purchasing decision. Nevertheless, the lack of accounting and costing data is evaluated as a barrier to this method being used commonly in the same publication.

Karande and Chakraborty (2013) used Measuring Attractiveness by a Categorical-Based Evaluation Technique (MACBETH) method for supplier selection. In this method, the evaluation of performance levels is done with semantic judgments of the decision maker. Values of the performance levels lie on lower and upper reference levels, which are associated with the values of 0 and 100, respectively. One of the main differences from AHP is that pairwise comparisons constitute linear differences here rather than ratios. One other important difference is that number of comparisons might be significantly lower than AHP since performance levels are compared rather than alternatives.

Min (1994) applied a Multi-Attribute Utility Theory (MAUT) to the supplier selection problem. In this study, 5 suppliers are evaluated with respect to the criteria set, and the utility of an alternative is found as the weighted sum of utilities from each criterion. Weights of the criteria are determined by the trade-offs between equally preferred alternatives. In other words, when two equally preferable alternatives have only two different attributes, the ratio between the differences in the attributes is considered the ratio of the priorities of the criteria.

Rezaei et al. (2016) used Best-Worst Method (BWM) for the supplier selection problem. BWM is an alternative MCDM method that still relies on pairwise comparison, but not all of the pairs. In this method, only the best (most important) and the worst (least important) alternatives (criteria) are compared to others. This method proposes to derive the weights with Non-Linear Programming as an MP technique, which minimizes the maximum deviation from comparison data (Rezaei, 2015). This nonlinear model is linearized in another paper by changing the error term (Rezaei, 2016).

2.3 Analytic Hierarchy Process

AHP is an MCDM method for evaluating a set of alternatives with respect to structured criteria set. AHP enables the decision-maker to make verbal judgments that are associated with numerical judgments. In this way, verbal judgments of decision-makers can be expressed as numbers. These judgments are obtained as a comparison of every criterion or alternative pair (Saaty, 1990). Then, the comparison matrix is formed by using the judgments, which are used for deriving the importance or performance levels. The transformation of linguistic terms to numerical scale is presented in Table 2.5 (Saaty, 1977).

Table 2.5: Scale of the Linguistic Terms in AHP (Saaty, 1977)

Verbal Judgments	**Associated Ratio**
Equal Importance	1
Weak Importance	3
Strong Importance	5
Demonstrated Importance	7
Absolute Importance	9
Intermediate Values	2, 4, 6, 8

A judgment matrix is formed as follows. A verbal judgment of the individual is taken, and its associated value is found in Table 2.5. This number is written in the row of the more important one and the column of the less important one. Its symmetric element is the reciprocal of the value obtained by the judgment. A value in the comparison matrix yields a ratio of the importance levels. To illustrate, say M is a comparison matrix, M(a,b)=x means that criterion a is x times more important than criterion b. For a weight vector derived from a consistent comparison matrix, these ratios are expected to be observed in the weights.

Varying levels of inconsistency are not rare cases in AHP. In fact, inconsistency might be observed due to several reasons even if the decision-maker is very consistent. One possible source of inconsistency is the integer scale of judgments. For example, if Criterion A is strongly more important than Criterion C, and Criterion C is weakly more important than Criterion B, then it is not possible to make a consistent comparison between Criteria A and C, since the associated ratio is 5/3 which does not exist in the scale. As another example, if Criterion B would be weakly more important than Criterion C, a consistent judgment between A and C would not be possible either since the associated value would be 15, which is beyond the upper limit of the scale. To conclude, the inconsistency term is inherent in AHP studied, and it should be dealt with appropriately.

Saaty (1977) also proposed a metric, namely C.I. (Consistency Index), to measure the inconsistency among the verbal judgments of the decision-maker. This metric depends on the largeness of the principal eigenvalue, whose minimum possible value is the number of elements to be compared. A higher deviation from its minimum value means a higher inconsistency. In order to measure this inconsistency independently from the number of elements, this difference from its minimum is normalized. At this point, it is lower the better for certain, but the question arises of how low is good?

One extreme example of inconsistency in the judgments would be a set of completely random judgments. When these random judgments and their reciprocals form a comparison matrix, it can be considered a reference point for inconsistency. Due to the randomness, various comparison matrices might be generated with varying levels of inconsistencies. For this reason, the average of C.I. values of these randomly generated matrices is defined as a reference point, Random Index (R.I.), depending on the number of elements to be compared. Saaty (1980) defines a metric Consistency Ratio (C.R.), which is the ratio between the C.I. values of the judgment matrix to be tested and the randomly generated judgment matrix. Afterward, an acceptable limit of inconsistency is determined based on this ratio. C.I. of the judgment matrix is considered acceptably consistent if it is less than or equal to 10 percent of the C.I. of the random index (R.I.).

$$C.I. = \frac{\lambda_{max} - n}{n - 1} \tag{2.1}$$

$$C.R. = \frac{C.I.}{R.I.} \tag{2.2}$$

AHP attracted great attention in various manners in terms of theory and application since it is presented. Harker and Vargas (1987) studied the performance of the 1-9 scale of AHP. In this study, five different scales are compared which are 1-5, 1-9, 1-15, x^2, and $\sqrt{x}$ scales. As a result, the 1-9 scale outperformed the other four scales. Beyond various studies to find a more representative general scale for AHP, Ahmed and Kilic (2022) studied the individualization of the scale with an experiment. Despite all the studies on the scale of comparison judgments for various profiles of decision-makers and problem settings, the 1-9 scale is still the most popular one in the AHP context.

There are several studies on deriving the priorities as well. For the comparison matrices which are not completely consistent, every element in the comparison matrix might imply a different ratio of the importance of the criteria. Originally, Saaty (1977) proposed the eigenvector method for deriving the weights. In this method, the weight vector is derived as the right eigenvector of the comparison matrix which is associated with the principal eigenvalue. Thanks to the pairwise comparison of all alternatives, this method performs well regardless of the scale of the criteria. Although this method is widely accepted, it has some drawbacks such as the obligation of a complete comparison matrix and combinatorial growth of the number of comparisons with respect to the number of alternatives.

Geometric mean method (Crawford, 1987) appears as an alternative way to the eigenvector method for deriving the weights which also relies on the pairwise comparison of all alternatives. One important difference of this method is the tolerability of missing data, i.e., incomplete comparison matrix. In this way, the weights can still be derived with fewer data but the desired number of comparisons is the same with the eigenvector method. In this paper, a distance expression (M) between two matrices is defined based on an error term as given in Equation 2.3, where A is the comparison matrix and C is a consistent matrix with the ratio of weights. The geometric mean method is proven to yield the m-closest consistent matrix to the comparison matrix.

$$m(A, C) = \left[\sum_{i} \sum_{j>i} (lna_{ij} - lnc_{ij})^2 \right]^{1/2} \tag{2.3}$$

With a similar approach, minimizing the sum of all error terms by an MP technique to derive the weights is proposed (Chandran et al, 2005). There are also computationally simpler approaches to derive the weights such as the row average of the normalized comparison matrix (Ishizaka and Labib, 2011). In this manner, it can be seen that even the same comparison matrix can result in different priority vectors with diversity in the approaches.

AHP is widely adopted in the industry for very different problems. Ho and Ma (2018) presented the industries as Manufacturing, Government, Logistics, Rental Cars, Food, Telecommunications, and more, in which AHP is used in integrated approaches.

2.3.1 Group Decision Making and AHP

One procurement staff can safely determine which supplier to place the order for a purchase requisition. Nevertheless, when the case is building a system rather than placing a single order, there are many employees affected by this system. Therefore, this system which provides recommendations should be enriched by the perspectives of diverse stakeholders for representativeness. In this sense, there is a set of decisions from different decision-makers which should be consolidated into one collective decision.

Group decision-making is not novel since the existence of multiple decision-makers is not a unique case. In an earlier study, Black (1948) studied a case that the decision-makers may have different preferences in a voting environment. In such a situation, there is no single action that is the best for every decision-maker.

Saaty's work (1994) which arises a debate on Arrow's Impossibility Theorem (Arrow, 1963) is promising in terms of combining different decisions into a group decision. The inference from the theorem is that a group decision cannot comply with all of the individuals all the time in terms of the ordering of the alternatives. One conclusion stated by Saaty is that there is not a single way to combine individuals' decisions, and different methods may result in an acceptable decision for all members. Moreover, it is stated that the alternatives can be associated with a cardinal scale rather than an ordinal scale. The Cardinal scale makes the problem gain more resolution to examine the satisfaction differences between the alternatives on an ordinal scale.

In the MCDM context, various methods can be used for group decision-making. Safarzadeh et al. (2018) proposed two revised mathematical models of BWM to estimate the optimal weights which comply with the individuals most. In this paper, the significance of the decision-makers is asymmetric, i.e., the weights of the decision-makers are different. Moreover, the objective functions of the models are different; which is an illustration of the claim that there are multiple acceptable ways to combine the judgments of the decision-makers. Yue (2011) proposed a solution by using TOPSIS for determining the weights of the decision-makers based on their representativeness of the average of the members.

Lai et al. (2002) conducted an AHP study for a software selection with the participation of six experienced individuals. When the judgments are analyzed in detail, it is seen that there exist some differences in the ordering of participants for both alternatives and criteria. These differences are not unexpected since the perspectives can be different. For instance, the supervisor of the group prioritized managerial factors. On the other hand, a group decision compatible with individuals is reached as the result of this study, which is the expectation from a group decision-making study.

In another selection study, Marcarelli and Squillante (2019) applied an AHP-based group decision-making method to select the best tender. In this study, multiple decision-makers are included in the second level of the criteria hierarchy just below the goal. Therefore, the contribution of the decision-makers to the goal in terms of weights might also differ; although it is equal in this particular publication. In the supplier selection topic, Sanayei et al. (2008) developed a mathematical model to maximize the total additive utility by using MAUT and LP. In this study, the utility obtained from allocating one unit of product to a supplier is determined as the weighted average of the utility levels of individuals for that supplier. In this way, the decisions of the members are combined into one average decision-maker, and the total utility of this average decision-maker is optimized under the constraints such as cost, quality, and capacity.

As it is stated in §2.3, judgments from the decision-maker are obtained as a comparison matrix which includes integers from 1 to 9, and priorities are derived from this matrix for a decision-maker. Forman and Peniwati (1998) studied the case that there are multiple decision-makers. It is stated that there are two ways to combine the decisions of individuals. One of them is Aggregating Individual Judgments (AIJ) and the other one is Aggregating Individual Priorities (AIP). The preference for using AIJ and AIP depends on the aim. If the group is preferred to act as a new individual, the judgments should be aggregated. On the other hand, if the individuals from the group are able to act according to their rights, then the priorities should be aggregated. In this publication, Aczel and Saaty's study (1983) is referred to with their finding that for satisfying unanimity and homogeneity conditions, individual judgments should be aggregated as the geometric mean of judgment matrices of the individuals. Besides, it is added by the authors that geometric mean should be used for this aggregation since the judgments in comparison matrices are ratio-based numbers. For the other way of aggregation, AIP, both arithmetic mean and geometric mean are reasonable ways. However, Escobar and Moreno-jiménez (2016) state that the geometric mean method is more commonly used, and it is the one that satisfies the unanimity condition, homogeneity condition, and reciprocal property.

In this section, a framework for group decision-making is provided from past studies in the literature. The discussion on the appropriateness of our case to the above-mentioned cases will be held in §4, and the adopted methodology will be explained with the motivations.

2.3.2 Decision-Making with Sparse Data and AHP

As stated in §2.3, the weights in AHP are derived from a comparison matrix with a size of nxn, where n is the number of elements to be compared. The decision-maker is asked to compare every pair of criteria. In this manner, the comparison matrix is expected to have only integers from 1 to 9, all associated with a verbal judgment. Nonetheless, there might be some obstacles to all of the pairwise comparisons, which constitutes a missingness in the comparison matrix.

Harker (1987) summarized possible reasons to have missingness in the comparisons. Two of the reasons are being unsure about the comparison of a pair and unwillingness to compare the pair. In both cases, the decision-maker would prefer not to answer these questions. The third reason is the big effort to make n(n-1)/2 comparisons when n is large. In the same publication, an approach for deriving the weights with missing data is proposed. The procedure is outlined as follows. Recall that for the complete comparison matrices, the weights are approximated by the principal eigenvector of the comparison matrix. Therefore the following equation holds, where C is the complete comparison matrix and **w** is the approximated eigenvector.

$$C\mathbf{w} = \lambda_{max}\mathbf{w} \tag{2.4}$$

In the case of missing data, when some elements of C are not known, the above equation can still hold if 0 is placed in the cell of missing data and 1 is added to the diagonal element of the related row. Further explanation for this approach is that since every column in C will result in w_i in each row of matrix multiplication, instead of an unknown element, this extra w_i can be obtained by the diagonal element. In this way, not only the equation will hold but also the comparison matrix will only contain real numbers rather than ratios of unknown weights. To conclude, the principal eigenvector approach is still applicable for the missing data case for the comparison matrix obtained by above mentioned procedure.

Saaty (2001) proposed a method for deriving priorities with less data. In this solution, the alternatives are ordered first, and then they are clustered. There will be principals in the clusters which exist in two different clusters. In this way, connectedness among the alternatives is assured, and priorities are derived within the clusters first, then global weights are calculated through these principal alternatives. With this approach, the classical 1-9 scale is extended as the upper limit may reach infinity. Ishizaka (2012) applied this solution for a supplier selection case and reduced the number of comparisons by 48.5%. Although this approach is promising for applying AHP with less effort, the requirement of ordering the alternatives might be a barrier to application, especially in a case there are multiple decision-makers.

BWM (Rezaei, 2015) can be preferred to reduce the number of comparisons. Because only the best and the worst alternatives are compared to others, the number of comparisons increases linearly with respect to the number of alternatives. Nevertheless, this method is not a solution for the cases that the decision-maker is unsure about or unwilling to make the comparison.

Oliva et al. (2017) proposed Sparse Eigenvector Method (SEM) for handling missing data in the comparison matrix. In this method, 0 is placed for the missing comparisons and diagonals in the comparison matrix as the first step. Then, every element in the comparison matrix is divided by the number of nonzero elements in their rows. In this way, the sparse comparison matrix is obtained. The idea behind this method is that in the case of a perfectly consistent comparison matrix, the principal eigenvector is preserved in the sparse comparison matrix. Therefore, the principal eigenvector of the sparse comparison matrix is an approximation for the priority vector in the sparsity case. In order to preserve the transition relation, there is an assumption of connectedness in this solution as well, which means every alternative should be reached by another at least through other alternatives. Oliva et al. (2018) conducted an experimental study using SEM and other approaches. In this publication, the preservation of the eigenvector is shown by an alteration in the eigenvector relation given in Equation 2.5. In complete case, this relation holds as in Equation 2.6. Required notations are provided beforehand. $\mathbf{w}$: Vector of Weights/Priorities

R : Complete and Perfectly Consistent Comparison Matrix

S : Sparse and Perfectly Consistent Comparison Matrix

D : Degree Matrix for S, D_{ij}=Number of nonzero elements in row i if i=j, otherwise 0

$$D^{-1}S\mathbf{w} = \mathbf{w} \tag{2.5}$$

$$R*\mathbf{w} = \begin{bmatrix} w_1/w_1 & w_1/w_2 & \ldots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \ldots & w_2/w_n \\ & \ldots & & \\ w_n/w_1 & w_n/w_2 & \ldots & w_n/w_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_n \end{bmatrix} = \begin{bmatrix} n*w_1 \\ n*w_2 \\ \ldots \\ n*w_n \end{bmatrix} = \lambda_{max}*\mathbf{w} = n*\mathbf{w} \tag{2.6}$$

If the above equations need an elaboration, in Equation 2.6, every multiplication for each element in the comparison matrix yields a result of w_i, and as a result of their sum, the right-hand side of the equation becomes $n * w_i$ for each row i. This balance does not hold for the sparsity case since the number of elements in each row may vary. Therefore, the rows of the sparse comparison matrix are divided by the number of nonzero elements to make the right-hand side equal to **w**. In this manner, $D^{-1}S$ is a matrix with a principal eigenvalue of 1, preserving the principal eigenvector of R.

A simpler alternative might be deriving the weights by geometric means of the rows (Crawford, 1987). If connectedness is assured, the priority of an element can be found as the geometric mean of the elements in its row. This method both tolerates the missing data, and multiple data for one comparison. Otherwise, the ratio of the priorities of all alternatives cannot be obtained from the comparison matrix. In addition, MP techniques such as Chandran et al. (2005) can also be used with their variations.

Chapter 3

PROBLEM DEFINITION

The organization whose supplier selection process is the subject of this study is an electronic company running operations in multiple business lines. Besides the core business field, this firm also needs diverse mechanical parts which are produced with machining processes. These mechanical parts have a big diversity in specifications such as geometric complexity, tolerances, and dimensions. In order to have good quality and the conformity of electronic and mechanical parts in higher-level systems, the organization carries out the design processes of the mechanical parts as well. Nonetheless, as a strategic decision, the organization does not prefer to have these production competencies to focus on its core business processes. Instead, it is preferred to outsource the production of these mechanical parts.

As a strategic decision, the organization prefers to work with a big number of suppliers rather than working with a few suppliers in big volumes. Two of the motivations for this strategic decision are the sustainability of its industry and having a more robust supply chain. Sustainability is assured by working with hundreds of suppliers of different sizes and by developing them throughout the supply process. In this manner, economic and operational sustainability, and consequently a robust supply chain is aimed. Although this objective sounds good, it is not easy for an organization to work with this many suppliers on complex production processes.

Even though outsourcing non-core business activities is an opportunity to focus on value-added activities, adding another firm to the supply process means taking the risks of the suppliers as well. In other words, the organization is not only responsible for its own operations but also for the operations of its suppliers. Arising from this responsibility and from the aim of sustainability, the organization works for having strong relationships with its suppliers. Therefore, the organization needs to recognize its suppliers extensively.

In this problem setting, procurement processes start with purchase requisitions. Since these requests arrive for the mechanical parts that are designed within the company, every mechanical part is unique, and no supplier markets the newly designed products beforehand. Moreover, there is no historical data for this mechanical part such as price or which supplier can produce. At this stage, the problem of the procurement staff is which supplier can produce this product with optimum cost and quality, and deliver on time. Hence, all the competent suppliers are candidates for supplying the part, and it is not possible to know which one of them will give the best quote with respect to the criteria before asking them. Moreover, even thinking of hundreds of suppliers is not easy at a high operational pace.

In the organization, purchase requisitions are shared by the staff according to the category of the products. Each procurement staff works with multiple firms and most of the suppliers work with multiple procurement staff from the organization for different product categories. One of the possible consequences of this structure is that the perception of a supplier might not be homogeneous within the organization. Another consequence is that a supplier might not be considered for some products if their responsible staff does not have knowledge of the supplier.

The supplier selection process starts with determining the competent suppliers. The organization may face several risks unless this decision is made correctly. One example of the risk is if a supplier is not sufficiently competent to produce the mechanical part and if it underestimates the complexity of the part, it might send a quote. Moreover, this quote might have a lower price than others due to the underestimation of the complexity, which can be misleading for the procurement staff. In such a scenario, one decision mistake can lead to various undesired consequences such as problems with quality, delivery, and additional costs.

Proceeding with the assumption that all the competent suppliers are listed successfully, a dilemma appears although having more quotes is favorable for the firm, it is costly and inefficient to receive quotes from all of the suppliers, especially in the case in which the number of suppliers is very large. Thus, the supplier selection process consists of two different stages which are determining the suppliers to be asked for a quote and deciding on which quote is the best.

The second stage mentioned above is not very challenging for the procurement staff since there are already a limited number of quotes, and the terms are defined with the quote. Thus, in this study the selection of the suppliers which are requested a quote for a mechanical part is focused, i.e., the purpose of this study is to select the best set of candidate suppliers among more than 400 suppliers with different qualification levels. The motivation here is to include the supplier in the shortlist which will appear as the best alternative in terms of the quote, without spending excessive effort.

To conclude, in the problem environment, the suppliers are desired to be matched with suitable mechanical parts with respect to their technical competencies. There is a two-way risk that makes this motivation arise. The first one is not being delivered on time with good quality because of the lack of competency of the supplier as discussed above. In this case, the deliveries of the firm might be affected and some extra costs may incur for the firm. The second risk is the probable increase in costs due to the unused competencies of the supplier. As a remark, the reason of the fact that the costs are focused on is that price information is unavailable at this stage because the purpose is to determine which suppliers to be asked for a quote. Moreover, as it is stated, procurement processes are initiated with purchase requisitions that arrive stochastically for a great variety of products. As a result of this variety and uncertainty, using an overqualified supplier for a product not only causes the second risk but also leads to the first kind of risk for other orders due to the unnecessary use of the capacity of the qualified supplier.

In line with the above discussion, the selection process of the set of suppliers depends on the complexity of the mechanical part. Thus, ranking the suppliers without considering the specifications of the mechanical part and selecting some of them does not work for this problem. In other words, the best set of candidate suppliers may significantly differ for an easy and difficult mechanical part. Consequently, an appropriate decision-making process should be sensitive to the specifications of the product to be sourced.

Chapter 4

SOLUTION METHODOLOGY

In this decision-making process, the purpose is to match the mechanical parts with a set of suppliers for an organization. Therefore, the solutions should be designed according to the requirements and constraints of the organization's problem. In general terms, the solution is desired to be explainable, comprehensive, and applicable.

The explainability term is related to the ease of understanding a result that is generated by the proposed solution, which is important for two main reasons. One of them is that the decision-maker should be able to see the trade-offs among the alternatives as their performances are often conflicting over criteria. Thus, procurement staff will be able to make their own decision in light of the analysis of the solution method. The second reason is that every supplier selection decision is questionable, and procurement staff want their decisions to be easy to explain. Explainability term is assured in this solution by using a widely used method with a reasonable level of mathematical complexity.

Comprehensiveness and applicability are two terms that a balance should be reached. The solution should be comprehensive for certain but covering all the cases and important points is not feasible due to the stochasticity of business processes. Being fully comprehensive requires having an infinite awareness of the system. For instance, even quitting a critical employee from a supplier might be an issue for the supply process; yet, computing the quitting probability of every employee from all the suppliers is not possible. This discussion brings us to the term applicability. This concept is handled as the organization's ability to evaluate a supplier within this scope. When the applicability is included in the big picture besides comprehensiveness, the question is on what variables our decisions should be shaped. In MCDM, this question points out the selection of criteria. Hence, these two terms are assured by selecting an appropriate set of criteria.

Once the criteria to be used are specified, evaluation of the suppliers' data with respect to them comes after. After evaluating the firm according to all criteria, the task to incorporate these performance measures is done by the selected MCDM method.

Last, a suitable subset selection method for the supplier list should be defined. In other words, in a list that every supplier is associated with numerical performance data, how many of them are to be included in the selection list is determined.

In light of these initial inputs about the problem and conducted literature research, 19 factors that are likely to affect the supplier selection process are determined. At this point, these factors are categorized so as to attain a good level of consistency. Relevant literature starts with Miller (1956) stating the existence of an upper limit for human beings in processing information. Saaty and Ozdemir (2003) examined this subject from the point of view of random inconsistency in the AHP context. As a result, they have found this limit valid and it is recommended not to exceed 7 for the sake of consistency. Therefore, 19 factors are categorized under 5 main criteria, which are Technical Competence, Quality, Environment and Safety, Delivery, and Finance. Moreover, the Technical Competence criterion is divided into two categories as Technical Infrastructure and Employees since it might be confusing to compare workers and machines.

Besides the aforementioned studies, different primitive solution approaches are developed. Nonetheless, the output of the studies until this step is only a draft study, which requires a deep discussion with the procurement staff who are the source of information. For this purpose, 3 different interviews with procurement managers are held for 1-2 hours each. In these interviews, both quantitative and qualitative evaluations of the managers are collected. The managers are asked to evaluate some statements. In this way, the accuracy of the initial findings on the subject will be analyzed. The question which is directed to the managers is as follows:

> ”Please state to what extent you agree with the statements below by using the 5-point scale. Note that 5 is the highest level of agreement and 1 is the highest level of disagreement. Also, provide an explanation for your evaluation.”

Collected answers from the participants are provided in the Table below. These statements and responses will be often referred to in the following subsections since the design of the solution is directly powered by this part of the study.

Table 4.1: Responses in the Interview with Procurement Managers

		Participants			
No.	Statement	1	2	3	Average
1	Suppliers can be evaluated with a general score that will be used for every procurement process.	1	4	2	2.33
2	Specifications of the mechanical parts should be taken into consideration in a decision support system for the supplier selection process.	4	4	5	4.33
3	Technical competence levels of the suppliers should always be evaluated as higher-the-better even if all suppliers are competent enough to produce the product to be ordered.	2	1	1	1.33
4	Criteria set which includes Technical Competence, Quality, Environment and Safety, Delivery, and Finance is required and sufficient for supplier selection.	4	4	2	3.33
5	Evaluation of technical infrastructure and employees is required and sufficient for evaluating the technical competence level of a supplier.	5	5	5	5
6	In the evaluation of the infrastructural technical competencies of the suppliers, it is necessary and sufficient to consider the machine infrastructure and special processes competency of the company as criteria.	4	5	5	4.67

Continued on next page

Table 4.1: Responses in the Interview with Procurement Managers (Continued)

		Participants			
No.	Statement	1	2	3	Average
7	In the evaluation of the employee-based technical competencies of the suppliers, it is necessary and sufficient to evaluate the number of employees of the company, the experience level of the production employees, the annual workforce turnover, the number of employees who completed the graduate education level related to the field of work and the number of employees who can use CAM.	4	3	2	3
8	Historical quality performance, quality control capabilities, and quality certificates are sufficient and required for evaluating the quality performance of the suppliers.	5	4	5	4.67
9	It is sufficient and necessary to evaluate the TRIR (Total Recordable Incident Rate), the quality certificates related to the subject, monthly electricity consumption, and the waste management policy in the evaluation of the suppliers on Environment and Safety.	3	4	3	3.33
10	To evaluate the suppliers according to the delivery criteria, it is sufficient and necessary to consider the percentage of your orders delivered on the promised date and the organization's share in the supplier's revenue.	4	4	4	4
11	In evaluating the financial reliability of the suppliers, it is sufficient to look at the firm's ability to find debt, the number of existing customers, and the cash ratio.	5	5	4	4.67

The remaining part of this section is organized as follows. In §4.1, the appropriate method will be proposed with a discussion of its pros and cons. In §4.2, the result of the criteria evaluation studies will be given. In §4.3, the method to transform supplier data to performance data concerning the criteria will be presented. In §4.4, the proposed method for preparing a shortlist of suppliers will be presented.

4.1 Method Selection

In this supplier selection study, an MCDM method will be used for evaluating the set of alternatives with respect to the criteria set. The importance of the method arises from the fact that the alternatives and criteria are put together as a system through this method. Therefore, upon determining the criteria set and evaluation procedures with respect to every criterion, the outputs of these separate judgments will be combined by the selected MCDM method.

Since this thesis focuses on a procurement process of a real organization, there are two practical concerns regarding the methods. The first concern on this subject is explainability as discussed in the previous section. To deal with this concern, the method should be easy to explain, so scores of the alternatives should be easily interpreted. In this way, the decision-maker will be able to see the trade-offs as well. For these motivations, additive utility methods seem to be ideal since the breakdown of total scores can be easily presented, and the alternatives can be compared with respect to the sub-criteria. The second concern is about the agility of the method. In other words, slight modifications should be easily made in order to comply with the changes in the dynamic business environment.

Supplier selection methods are categorized by de Boer et al. (2001) into five. The first decision model is the linear weighting model in which the criteria are weighted with respect to their importance levels. Then, alternatives are evaluated with a total score of the weighted scores gained from the criteria. The second is the Total cost of ownership (TCO) model, in which all quantifiable costs are considered. In this manner, it has a wider perspective compared to the price, yet expressing all the costs in numbers is not easy. For instance, the cost of a delay in some delivery might not be easily calculated since there are various consequences. The other three are Mathematical Programming models, Statistical Models, and Artificial Intelligence (AI) models. These three sophisticated models might be promising in such complex problems. However, even if these models are successfully applied, interpretation of the results is demanding in terms of workforce competence. Thus, these methods will not be preferred for this problem since they contradict the explainability concern.

Based on the inputs from the procurement staff and the literature review studies presented in this section and §2.2, it is determined to select a method that is widely used. Moreover, among several others, using a linear weighting model is determined to improve the explainability of the outputs. After the set of alternatives is narrowed down to commonly used linear weighting models, two review articles are studied to see the comparison of the alternative methods.

Chai et al. (2013) reviewed 123 articles on the supplier selection topic. According to the article pool, AHP is the most frequently used method for supplier selection problems. Although the frequency of the usage of this method decreased in the next years according to the article pool in Chai and Ngai (2020), it still appears as the most commonly used MCDM method for supplier selection. According to these two publications, AHP is followed by Analytic Network Process (ANP) and Technique for Order Performance by Similarity to Ideal Solution (TOPSIS) in terms of frequency. The other MCDM methods reviewed in both of these articles are PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluation), VIKOR, and DEMATEL (Decision Making Trial and Evaluation Laboratory). Afterward, the frequency of usage of these methods in the supplier selection context is tried to be explored. Nevertheless, the set of articles in which the names of these methods occur is too big to be reviewed. Therefore, a more shallow analysis is conducted by the number of articles.

For this research, Google Scholar is used as the database of the articles. In order to find related articles, three different filters are used. The first filter is to include the exact phrase of ”supplier selection”. Secondly, the ”include citations” option is deactivated since unavailable articles cannot be confirmed. The last filter is the name and abbreviations of the methods. The filter is designed to return the results which include either the abbreviation or the full name of the method in the title. Because many articles refer to almost all methods in the Literature Review section, these words are searched in the title of the articles. It should be noted that the outputs for different filters have intersections such as the articles using both AHP and TOPSIS for supplier selection. This search is made at the end of June 2023. The results of these searches are presented in Figure 4.1.

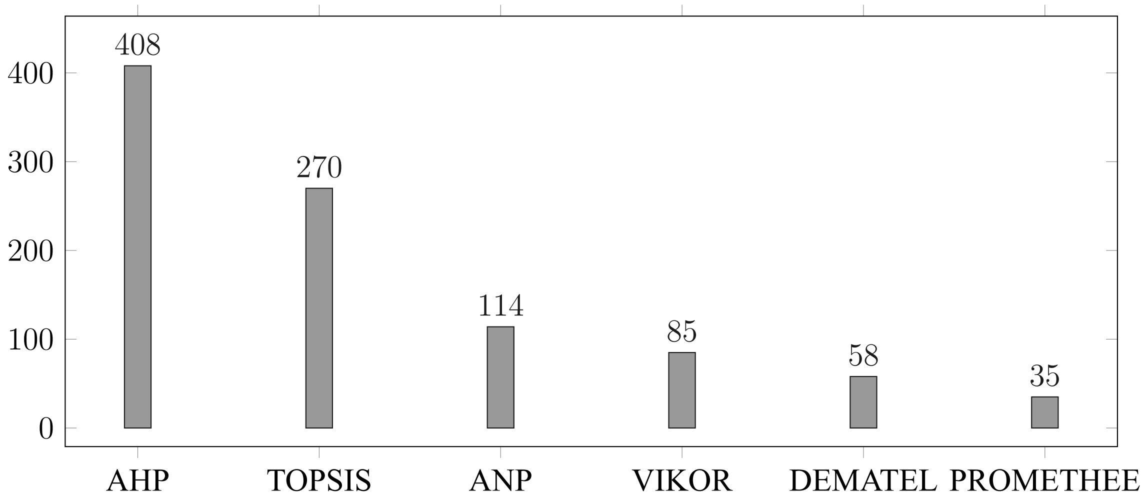

Figure 4.1: Number of Results Regarding MCDM Methods

According to the results presented in the figure above AHP seems to be the most frequently observed MCDM method in supplier selection context. Besides, there are several advantages of using AHP. The first advantage is that since it is a linear weighting model, it is very easy to show the breakdown of the total score. In this way, a procurement staff will be able to see the comparisons of the alternatives with respect to sub-criteria as well. Moreover, they will be able to explain their decisions which are supported by this system with quantitative arguments. The second advantage is that while these methods are briefly presented to the staff, AHP is known by some of the participants, yet they did not know other methods. Therefore, the users might feel more familiar while using AHP, and there is some staff who can manage the slight changes in the model. Last, AHP is commonly used in the integrated models as discussed in §2.2, which means some variations can be made on the model based on AHP if needed. Accordingly, using AHP is found to be useful, but there will be a need for variation for evaluating the firms since there are too many suppliers to make pairwise comparisons.

To conclude, it is determined to decide on the priorities of the criteria by using AHP. Thanks to the common usage and an acceptable level of complexity of AHP, the organization will be able to change the criteria set and priorities over time. Nevertheless, the classical hierarchy structure of AHP which includes the alternatives at the bottom level cannot be used due to the big number of alternatives and a big number of constantly arriving purchase requisitions. Therefore, the evaluation methods will be tailor-made to the organization which is shaped by the literature and group interviews with the procurement staff. The details of the evaluation procedure are presented in §4.3.

4.2 Criteria Selection

The criteria in an MCDM problem indicate what to consider while making a decision. In this sense, it is the first place where a decision begins to form beyond all the mathematical operations. For this reason, a serious effort has been put into this subject both in terms of time during the studies and in terms of pages in this thesis to give the deserved importance.

The information collected in the survey will be consulted while deciding on the final set of criteria for supplier selection. Statements 4-11 in the survey are directly related to the suitability of the criteria. The average agreement points of the three participants to the statements will be a guide to determine what to do with the criteria given in Table 4.1. There are three possible actions which are elaborating or narrowing down the scope of the criteria, and using as in the statement. If there is a consensus on the lack of comprehensiveness of a criterion, then more details should be included. On the other hand, if there are concerns concerning the applicability objective for a criterion, then it should be revised to a leaner version. If a criterion is sufficiently comprehensive and also not cause a risk in terms of applicability, then it is decided to use the criterion as it is.

Statement 4 defines the general framework of the supplier selection process based on the initial findings. For this statement, all of the participants stated that the communication levels of the suppliers must be included in this system because the complexity of the procurement process is high. In other words, there might appear various problems throughout the road, and they should rely on the collaboration of their suppliers. Hence, it is determined to add the sixth criterion ”Communication” after these interviews. Besides, Participant 3 declared that the attitude of the supplier toward the environment and safety is a more strategic objective, and it should not directly affect an operational-level decision. However, this criterion is not removed due to the increasing importance of these subjects and the agreement of two other participants.

Examination of the technical competence of the suppliers under two subcategories is approved by all of the participants. There is neither a request for elaboration nor for dividing into these categories. Therefore, this branching is found appropriate, and it will be used as it is declared in Statement 5.

For Statement 6, Participants 2 and 3 declared that they fully agreed and had nothing to add. However, Participant 1 had concerns about the sufficiency of these two criteria in determining the level of infrastructure technical competence. For example, he stated that the software used by the supplier is also important. Although there are other factors that can be considered important, it is not considered operationally feasible by Participant 1, too, to monitor the suppliers according to these criteria. Thus, in the end, a new criterion was not proposed and it was decided to use these two criteria as they are.

Purchasing managers were more concerned about assessing employee-based technical competencies. Although they state that this criterion is one of the most important ones since good personnel can produce good results, it is almost impossible to monitor and evaluate this personal data. Therefore, the system should depend on the declaration of suppliers, which can be risky for this large set of data. Because of these concerns, agreement levels are at a medium level on average for Statement 7. However, the participants stated that this evaluation could work reasonably well if the criterion of ”the number of employees who completed the postgraduate education” was removed and the ”number of employees” was revised as ”the number of employees in production”.

Participants 1 and 3 completely agreed with Statement 8, and they prefer to use this branching for assessing the supplier in terms of quality criterion. Nonetheless, Participant 2 had reservations about the reflection of the quality certificates on the business processes. Thus, the quantitative answer here was 4, but after a short discussion, he agreed to keep this criterion because having a certificate points out an awareness even if the way that suppliers do business does not completely comply with how it is defined in the certificate.

In Statement 9, the participants stated that if a supplier does not fulfill its responsibilities in terms of environment and safety, working with that supplier is already rejected from the very beginning as a strategic decision. Therefore, they still had hesitancy about the existence and evaluation of this criterion. Moreover, ”monthly electricity consumption” is not considered appropriate. The reasoning here is that despite the fact that efficiency in using energy is very important, total consumption highly depends on the output which differs from one supplier to another. Yet, they agreed to include this criterion in the end, if electricity consumption is excluded.

For assessing the suppliers with respect to delivery criteria, all of the attendees agreed to use historical timely delivery performance as a metric, but they stated that it should not be a binary evaluation. In other words, being late for 1 day and 120 days are not equivalent since the longer the supplier is late, the greater the loss undergone by the organization. Thus, this criterion is revised as ”historical delivery performance” as a general expression, and the evaluation procedure will be elaborated in the next section.

In the last question of the survey, the evaluation of the financial reliability of the suppliers is attempted to explore. In this criterion, the purpose is to determine whether a supplier is able to manage the financial process and produce the mechanical part. For this motivation, cash ratio, number of existing customers, and ability to find a loan are found useful. Participant 3 hesitated about loans since the suppliers are generally expected to manage the process with their own resources. However, after a discussion, the participants have reached a consensus on the fact that using a loan is not a problem, and it can be interpreted that the firm is loyal to the debts if the firm can find a loan. In this instance, this criterion is decided to be used as it is.

As the result of this study and the above discussions, the criteria hierarchy is finalized with 18 lowest-level criteria of which Communication is the main criterion. The motivation for not adding sub-criteria to this criterion is that the procurement managers explicitly stated that this criterion should reflect the subjectivity of decision-makers. In other words, this criterion is expected to reflect the opinions of the procurement staff without being framed by metrics. Finalized hierarchy of criteria is given in Figure 4.2.

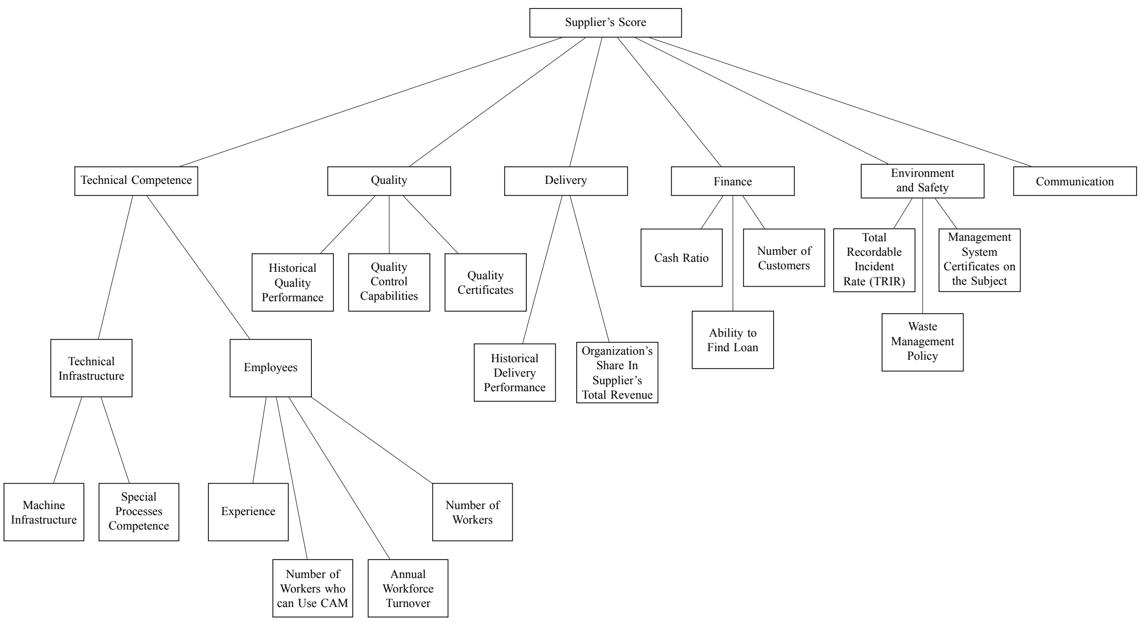

Figure 4.2: Hierarchy of the Criteria for Supplier Selection

It should be reminded that the purpose of this study is to build a decision support system for determining the suppliers to be asked for a quote for a mechanical part. Therefore, this system should be enriched by not only the literature but also the users of the system. Accordingly, most of the criteria in the finalized hierarchy are neither unique nor novel since they are commonly used in the literature and they have an impact on the operations, which is a desired output. Although related articles are reviewed in §2.1, only the subsets of the criteria set which are prioritized are presented earlier to save space. Thus, there were many others in related publications. Hence, upon finalizing the criteria hierarchy, the existence of each criterion in the reviewed articles is reported in Table 4.2. Despite the fact that different authors use different wordings, criteria are reported to exist in a publication if a criterion in the article exists with a very similar scope.

Table 4.2: Usage of the Selected Criteria in the Articles

		Articles						
Top Level Criteria	**Bottom Level Criteria**	**Sarkis and Talluri (2002)**	**Huang and Keskar (2007)**	**Verma and Pullman (1998)**	**Barbarosoglu and Yazgac (1997)**	**Kannan and Tan (2002)**	**Chan and Chan (2004)**	**Rezaei et al. (2016)**
Technical Competence	**Machine Infrastructure**	X	X		X	X	X	X
	Special Processes Competence		X			X		
	Experience of Workers					X		
	Number of Workers who can Use CAM				X			
	Annual Workforce Turnover		X					
	Number of Workers		X		X	X		
Quality	**Historical Quality Performance**	X	X	X	X	X	X	
	Quality Control Capabilities				X	X		
	Quality Certificates	X	X		X			X
Delivery	**Historical Delivery Performance**		X	X	X	X	X	
	Organization's Share in Supplier's Revenue				X	X		
Finance	**Cash Ratio**					X		
	Ability to Find Loan							
	Number of Customers							
Environment and Safety	**Total Recordable Incident Rate**		X					
	Waste Management Policy							X
	Management System Certificates on the Subject		X					X
Communication	**Communication**	X	X	X		X	X	

Observing the table above, it can be seen that all the criteria but the criteria Ability to Find Loan and Number of Customers are handled in the reviewed articles. One guess for not observing these two criteria is that there are various ways to evaluate financial strength. Although there were financial criteria in almost all of the it is observed that sub-criteria are not very similar among the articles. For instance, the ability to find a loan is evaluated in this study since the supplier needs to purchase a big amount of raw material at once; however, this criterion might be meaningless in continuous production environments since the rate of income and expenses are less fluctuating. Therefore, financial stability might be more important in such a case.

One interesting observation is that the criteria concerning the environment and occupational safety are not commonly observed in the earlier publications in the reviewed literature. Besides, technical-infrastructure-related criteria are more common than employee-based criteria. Last, the historical performance of the suppliers in different categories such as delivery and quality is considered predictive of future performance and is included frequently. To conclude, inputs from the procurement staff and the literature formed a criteria hierarchy with varying levels of similarity to the existing literature. In this way, both the existing literature is utilized and also tailor-made criteria set is defined for the organization.

Up to this point, the purpose was to understand the need for changes in the criteria set. Hence, only the criteria for which participants have reservations about or objections to them are discussed. Upon determination of the criteria set and building the hierarchy, every criterion is explained with associated motivations as follows in the next paragraphs.

Technical Competence: The technical competencies of the suppliers are the parameters that show the manufacturability of the ordered parts within the supplier company. Thanks to these parameters, suppliers will be able to be evaluated according to their potential. Relevant competencies are examined in two lower-level sub-criteria which are Technical Infrastructure and Employees.

Technical Infrastructure: The scope of this criterion is the physical and technical qualifications and infrastructure of the supplier companies. In this criterion, the machining infrastructure such as CNC (Computerized Numerical Control) Cutting Machines, and special processes competence such as painting, coating, and heat treatment are evaluated.

Machining Infrastructure: Machining Infrastructure is one of the most important criteria that determine whether a company can produce a mechanical part or not. The number of axes, axis limits, maximum spindle speed, and many other features of CNC Machines will be decisive in the evaluation of suppliers in terms of machine infrastructure.

Special Processes Competence: In the case that the mechanical part requires special processes, there are certain supplier candidates that have the ability to run these processes within their firms. Although these special processes can also be outsourced by the suppliers as well, the firms that have competence in these processes are preferred for two main reasons. One of them is the complexity arising from adding another firm into the process. Another one is that the mechanical part can be damaged during multiple times of transportation. Therefore, the existence of the competence of the suppliers in these special processes is evaluated.

Employees: The workforce of the suppliers can be evaluated in many aspects such as numerical, quality, and continuity and it is of great importance from the beginning to the end of the supply process. Employee-based competencies of the firms are evaluated with 4 sub-criteria under this criterion.

Average Experience of the Workers: Besides the machine infrastructure, the know-how of the workers also has a critical impact on the manufacturing process. For instance, if a mechanical part has a difficult geometry and tolerances, past manufacturing experiences of the workers will gain more importance since a technician who produced a similar part before knows the solutions to various problems beforehand in the manufacturing process.

Number of Workers: The supplier of a mechanical part is not only the manufacturer of the product. Instead, this firm is the procurer of the raw material, the planner of the production process, the finance responsible for business processes, and more. Therefore, in addition to the technical personnel who are direct laborers, there are also other staff who will contribute to the supply process by resolving problems arising from process complexity. Hence, the suppliers with more human resources can be considered candidates for more complex mechanical parts.

Number of Workers who Can Use CAM: Employees who can use CAM (Computer Aided Manufacturing) software are the human resources that produce the mechanical part of the machining process. As a result, an increase in this number means more and diverse cumulative knowledge and experience of manufacturing processes, enabling firms to produce more complex mechanical parts.

Annual Workforce Turnover: As mentioned in Criterion 3, know-how and experience have a critical impact on the process in several manners. One is the efficiency arising from the learning curve. Another one is more accurate planning due to past experiences with the manufacturing processes of similar products. At this point, keeping this experience and know-how within the firm is crucial since the knowledge is not stored by the firms. Instead, workers store this knowledge, and losing the workers means losing the know-how. Therefore, it is preferred for the suppliers to keep the workforce turnover low.

Quality: A complex electronic system is composed of a great number of electronic components, mechanical parts, and lower-level systems. Since the system is designed as a whole, its sub-parts should meet the requirements for the system to work conveniently. In this sense, suppliers should be dependable that they supply the products which fulfill the specifications in the design of the parts. Under this main criterion, the historical performance of the suppliers, quality control capabilities, and quality management certificates are evaluated.

Historical Quality Performance: The definition of undamaged and fault-free products covers the products that the supplier company successfully carries out all the production, transportation, and purchasing processes and delivers ready for use. If the historical quality performance of a supplier is good, it is predicted that this firm will also supply well quality products in the future. Hence, firms with good historical quality performance should be preferred in the next orders.

Quality Control Capabilities: With a more result-oriented approach, the firms can be evaluated in terms of quality criterion with their historical performances only. However, this approach might be too superficial because of the diversity of the mechanical parts. In other words, a firm can supply easy parts many times, but it does not mean that this firm is able to produce a more complex mechanical part defect free. Thus, their quality control capabilities should also be considered, and the Coordinate Measuring Machine (CMM) infrastructure of the firms will be evaluated for this purpose.

Quality Certificates: Having a good quality control mechanism saves the firm from supplying bad quality products, but it does not mean the firm does not produce waste as a product. For this purpose, it is required to create quality at the source, which is possible by having certain standards in business processes. The existence of these standards will be evaluated with the existence of quality management certificates in the firm, and the firms having these certificates will be considered that they have audited good standards in their business processes.

Delivery: The plans of the organization are based on the deliveries from the suppliers. In this situation, working with suppliers which align their deliveries with the organization's demand saves invisible costs. Some examples of these costs are the cost of idleness of workers and machines, compensation due to the delay of the deliveries to end users, and inventory holding costs for other materials which are delivered on time. Hence, delivery is a very important criterion, and it will be evaluated by the sub-criteria Historical Delivery Performance and Organization's Share in the Supplier's Total Revenue.

Historical Delivery Performance: In this criterion, the suppliers are evaluated with their deliveries in the past to the organization. As stated earlier, the organization which is the subject of this study is an electronic company, and it produces higher-level systems by using mechanical parts. Therefore, the delivery of a mechanical part also precedes the organization's operations. Hence, receiving the deliveries on time is crucial for the organization. Similarly to Historical Quality Performance, it is predicted that a supplier with successful deliveries in the past are going to make successful deliveries in the future. Thus, suppliers who make their deliveries on time or with small deviations from the commit date are preferred over others.

Organization's Share in the Supplier's Total Revenue: The organization in this study runs tens of projects at the same time of which each has high process complexity and uncertainty. For this reason, the plans of the organization are subject to change due to various reasons. In this case, when production plans and priorities change, procurement of some parts might gain urgency. Although the suppliers are only responsible for the committed date, working with a supplier who is willing to align their operations with the organization might be effective in resolving possible problems of the organization. Therefore, it is preferred to work with the suppliers of which the organization has a significant portion in their revenues since it is considered that the organization's demands are given more importance by these suppliers.

Finance: Every step of the supply process from sourcing to delivery results in several financial transactions. In this sense, suppliers are dependable only if they are capable of managing these transactions to have a solid financial status. Moreover, the organization which is the subject of this study cares about the sustainability of its supply chain stakeholders. Therefore, the suppliers are not expected to supply the products at a loss. Instead, the expectation from them is to maintain their businesses. Hence, the suppliers are evaluated for their financial success with sub-criteria Cash Ratio, Ability to Find Loan, and Number of Customers they have.

Cash Ratio: The suppliers are desired to run their business processes and manufacture the mechanical parts with their own resources. However, the manufacturing process leads to several expenses such as labor costs, material costs, facility expenses, and so on. In this situation, the organization wants to rely on the suppliers' financial ability to run their business processes. In other words, sufficient financial strength to cover the expenses is expected. For this purpose, suppliers' ability to cover their debts and have a remaining resource is evaluated by the cash ratio.

Ability to Find Loan: A firm might be able to run the supply process even if they do not have enough cash, by finding monetary resources by taking a loan. Because finding a foundation that is willing to loan requires good financial performance and loyalty to debts in the past, this method is not harmful to the reliability of the supplier. In Turkey, Kredi Kayıt Bürosu (KKB), which is founded by 9 leading banks, developed a widely accepted financial measure, namely Findeks, which provides an idea of the risk of not being paid back. As a result, the suppliers with higher Findeks ratings are considered competent to find necessary resources and they are preferred over other candidates.

Number of Customers: Customers are sources that the revenue of the organization is generated. In this sense, having more customers enables firms to run their businesses without being affected by the demand and payment fluctuations of their customers. In this manner, suppliers with a higher number of customers will have more balanced cash flows. Hence, they are considered to have a more robust financial structure, which is important for their dependability.

Environment and Safety: Every company has various social responsibilities besides its commercial responsibilities to the planet, its employees, and its customers. The attitude of the suppliers towards these responsibilities is an indication of the respect that the company offers to its environment, society, and directly related stakeholders such as customers. Therefore, in the evaluation of suppliers, the sensitivity and actions of the suppliers regarding these responsibilities are included.

Total Recordable Incident Rate: In this criterion, the sensitivity of the supplier to occupational health and safety is evaluated. TRIR is a metric obtained by finding the projection of the recordable occupational accident rate that occurs within the company to a company that works 40 hours a week and 50 weeks a year for 100 people. Thanks to this measure, companies with different sizes can be evaluated on the same scale. Suppliers will be given a score with respect to this ratio, and firms with a lower ratio will be considered preferred supplier candidates since they pay attention to the safety of their workers.

Waste Management Policy: In machining processes, a big amount of waste might be generated. For instance, the creation of dust in machining processes is unavoidable because the nature of the process requires removing some part of the material and shaping the part to its final form. Not limited to dust which is created from a scarce resource, there are many types of waste such as facility waste, lubricants, defective products not being able to be reworked, and so on. Thus, it is the responsibility of every firm to manage these wastes in a respectful manner to nature. Therefore, the existence of a well-structured waste management policy is included in the evaluation process of the suppliers.

Management System Certificates on the Subject: The firms with certificates of management system standards are considered to have an awareness that is translated into business processes since the certificates are given upon an audit process. These standards are appreciated by the organization, and the firms with those certificates are preferred over others.

Communication: This criterion is both a main criterion and also one of the lowest-level criteria. The reason for not dividing this criterion into lower-level criteria is that procurement staff has a consensus on that the evaluation of communication should be made directly by subjective judgments of them. In such a complex procurement environment, communication is key to resolving various problems beforehand. For instance, a supplier might not be able to make a delivery on time, yet informing the organization about the current situation is still valuable for enabling the organization to revise the plans accordingly. As another example, when there are small defects in the product, the firm should be willing to run rework operations and work collaboratively to prevent further mistakes in the process. Therefore, the procurement staff will evaluate the suppliers' communication success, and the suppliers with higher stakeholder satisfaction will be preferred over others.

After the criteria set is determined and the hierarchy is built, an AHP study is conducted with 15 participants in total. This 15-people participant group is composed of 9 Procurement Engineers, 3 Procurement Managers, 2 Supply Chain Strategy Engineers from the organization, and 1 Professor studying in this field. Before receiving the judgments of the participants, they are informed about AHP and the hierarchy of the criteria. In this way, most of the participants could contribute with consistent judgments without any confusion about the scope of the criteria.

For each participant, 8 comparison matrices are formed. In the first matrix, M1, six main criteria are compared. In Matrix M2, two sub-criterion of Technical Competence are compared, which yields a consistency index (C.I.) of 0 since there is no source of inconsistency in comparing the 2 criteria. A similar result appears in Matrix M3 in which sub-criteria of Technical Infrastructure are compared. For the comparison matrices with two elements, C.R. cannot be calculated since the random index (R.I.) is also equal to zero. These matrices will be considered consistent for all of the individuals. In Matrix M4, the sub-criterion of ”Employees” sub-criterion are compared. In the Matrices of M5, M6, M7, and M8, the sub-criteria of the Quality, Delivery, Finance, and Environment and Safety main criteria are compared, respectively. In Matrix M6, no source of inconsistency exists since there are two sub-criteria of the main criteria ”Delivery”, similar to M2 and M3. The consistency ratios of the participants for the comparisons with more than two elements in these matrices are given in Table 4.3.

Table 4.3: Consistency Ratio (C.R.) Values for the Comparisons in Matrices 1-8

	Comparison Matrices				
Participant Number	**M1**	**M4**	**M5**	**M7**	**M8**
1	0.189	0.064	0.056	0.010	0.170
2	0.090	0.057	0.040	0.040	0.040
3	0.249	0.096	0.523	0.152	0.757
4	0.171	0.226	0.543	0.308	0.089
5	0.139	0.073	0.141	0.010	0.006
6	0.034	0.050	0.077	0.000	0.000
7	0.048	0.012	0.077	0.000	0.000
8	0.102	0.097	0.141	0.030	0.141
9	0.009	0.156	0.000	0.000	0.000
10	0.205	0.127	0.056	0.000	0.019
11	0.071	0.122	0.056	0.141	0.056
12	0.077	0.087	0.056	0.000	0.000
13	0.063	0.097	0.000	0.04	0.000
14	0.054	0.006	0.077	0.077	0.077
15	0.058	0.098	0.003	0.000	0.000

The AHP study for evaluation of the criteria is conducted one on one after general information about AHP is provided to the participants. Therefore, the participants made their judgments in an interactive environment, and they have taken feedback on their judgments if there is a high inconsistency among the pairwise comparisons. In this sense, it is known that the participants tried to be consistent. Nevertheless, there still exists some unacceptable Consistency Ratio (C.R.) values. The threshold for C.R. is determined as 0.10 as in Saaty (1980).

When Table 4.3 is observed, it can be seen that 8 of the participants have at least one C.R. value greater than the threshold. Although they have made some inconsistent judgments, this observation will not be generalized because the inputs from these experts are valuable. In other words, the matrices with a C.R. value less than 0.10 of these 8 participants will not be excluded from the calculations. For instance, Participant 8 will not contribute to the evaluation of the main criteria; however, the judgments about the sub-criteria of the Finance will be used.

As it is mentioned in §2.3.1., there is not a single correct way for aggregating the decisions of the individuals. Moreover, this decision should be made according to the problem environment. In this context, there are multiple procurement staff who purchase similar items from diverse suppliers. Since they are different human beings, they have different perspectives, focuses, and backgrounds. Nevertheless, they all work for the same organization which has a single set of objectives. Therefore, they are expected to act as a whole in their decisions. In this manner, their decisions can comply with the organization and each other. Thus, individuals' decisions are aggregated on a judgment basis rather than priorities. For Aggregating Individual Judgments (AIJ), considering the geometric mean of individual judgment matrices as a judgment matrix of a new individual is widely accepted, and this method is preferred in this study. The resulting judgment matrices are provided in Tables 4.4-4.11.

Once the new judgment matrices are obtained from sufficiently consistent comparison matrices of the individuals, the priorities are derived with the right principal eigenvector method. In line with this adopted methodology, C.R. values are reported in Table 4.12 with respect to the associated Tables 4.4-4.11. Besides, derived priorities of the main criteria and sub-criteria are presented in Table 4.13.

Table 4.4: Comparison Matrix (C.M.) of Six Main Criteria and Abbreviations

	C1	**C2**	**C3**	**C4**	**C5**	**C6**
C1	1	0.485	0.624	2.509	5.586	1.023
C2	2.061	1	0.806	2.868	5.654	1.434
C3	1.603	1.241	1	3.100	5.949	1.767
C4	0.399	0.349	0.323	1	3.281	0.432
C5	0.179	0.177	0.168	0.305	1	0.192
C6	0.978	0.697	0.566	2.312	5.211	1

Criteria List
C1: Technical Competence
C2: Quality
C3: Delivery
C4: Finance
C5: Environment and Safety
C6: Communication

Table 4.5: C.M. of the Sub-Criteria of Technical Competence and Abbreviations

	C1	**C2**
C1	1	1.013
C2	0.987	1

Criteria List
C1: Technical Infrastructure
C2: Employees

Table 4.6: C.M. of the Sub-Criteria of Technical Infrastructure and Abbreviations

	C1	**C2**
C1	1	1.441
C2	0.694	1

Criteria List
C1: Machine Infrastructure
C2: Special Processes Competence

Table 4.7: C.M. of the Sub-Criteria of Employees and Abbreviations

	C1	**C2**	**C3**	**C4**
C1	1	2.927	3.037	3.819
C2	0.342	1	0.833	1.944
C3	0.329	1.201	1	1.63
C4	0.262	0.514	0.613	1

Criteria List
C1: Experience
C2: Number of Workers who can Use CAM
C3: Annual Workforce Turnover
C4: Number of Workers

Table 4.8: C.M. of the Sub-Criteria of Quality and Abbreviations

	C1	**C2**	**C3**
C1	1	2.868	3.89
C2	0.349	1	2.408
C3	0.257	0.415	1

Criteria List
C1: Historical Quality Performance
C2: Quality Control Capabilities
C3: Quality Certificates

Table 4.9: C.M. of the Sub-Criteria of Delivery and Abbreviations

	C1	**C2**
C1	1	2.564
C2	0.39	1

Criteria List
C1: Historical Delivery Performance
C2: Organization's Share in Supplier's Revenue

Table 4.10: C.M. of the Sub-Criteria of Finance and Abbreviations

	C1	**C2**	**C3**
C1	1	1.272	0.797
C2	0.786	1	0.675
C3	1.254	1.482	1

Criteria List
C1: Cash Ratio
C2: Ability to Find Loan
C3: Number of Customers

Table 4.11: C.M. of the Sub-Criteria of Environment and Safety, and Abbreviations

	C1	**C2**	**C3**
C1	1	2.509	0.59
C2	0.399	1	0.313
C3	1.696	3.196	1

Criteria List
C1: Total Recordable Incident Rate
C2: Waste Management Policy
C3: Management System Certificates on the Subject

Table 4.12: Consistency Ratio (C.R.) Values for the Aggregated Judgment Matrix

Table Name	Table 4.4	Table 4.7	Table 4.8	Table 4.10	Table 4.11
C.R.	0.062	0.028	0.064	0.000	0.002

Table 4.13: Global Weights of the Criteria for the Aggregated Judgment Matrices

Criteria	**Global Weight**
Technical Competence	**0.176**
Technical Infrastructure	0.089
Machine Infrastructure	*0.052*
Special Processes Competence	*0.036*
Employees	0.087
Experience	*0.045*
Number of Workers who can Use CAM	*0.016*
Annual Workforce Turnover	*0.017*
Number of Workers	*0.010*
Quality	**0.253**
Historical Quality Performance	0.155
Quality Control Capabilities	0.065
Quality Certificates	0.033
Delivery	**0.273**
Historical Delivery Performance	0.196
Organization's Share in Supplier's Revenue	0.077
Finance	**0.087**
Cash Ratio	0.029
Ability to Find Loan	0.023
Number of Customers	0.035
Environment and Safety	**0.036**
Total Recordable Incident Rate	0.012
Waste Management Policy	0.005
Management System Certificates on the Subject	0.019
Communication	**0.175**

4.3 Evaluation of the Alternatives

Value functions are used to determine the satisfaction level from the performances of the alternatives concerning a criterion for a decision-maker. Typically, in AHP there is no need to worry about the value function as every pair of criteria is compared, and the value that is obtained from the alternatives is determined. Although this approach is widely accepted in the literature, comparing every pair of alternatives with all criteria every time a mechanical part is ordered is time-consuming and sometimes infeasible. For instance, when tens of procurement requests arrive in a day, comparing every pair of alternatives might be time-consuming. As another example, when there are too many alternatives, consistency among a big number of judgments is suspicious. Therefore, it can be preferred to determine the values of the alternatives by using a predefined procedure.

Stevens (1946) categorized the scales into four groups, which are nominal, ordinal, interval, and ratio. In this categorization, the common point of interval and ratio scales is that they are both quantitative, but the ratio scale has a true zero such as the number of workers. In an interval scale, it is not possible to state a true zero such as delivery performance. For instance, a firm might say a delay of more than 5 days is not acceptable, yet having a delay of 6 days cannot be considered a true zero. Besides, 2 days of delay may not be twice as bad as a 1-day delay. The other two scales are categorical scales in common, but the ranks of the alternatives are known in the ordinal scale. For example, a high-honor student has a higher GPA than an honor student, which is a demonstration of the ordinal scale. Nonetheless, this interpretation cannot be done for some criteria such as eye colors. In light of these discussions, it is clear that the rankings of the alternatives are known except for the nominal scale. Since the evaluation of the alternatives is not completely ambiguous for these 3 scale types, other methods than pairwise comparison could be useful by utilizing the scale type of the performances.

Although it is known that pairwise comparison has some drawbacks, it handles several risks and obtains the priorities because the derived priorities directly imitate the preferences of the individuals through their judgments. On the other hand, the values obtained from the alternatives might be modeled if individuals' way of making a decision is explored. With this motivation, in the first three questions of the survey, the approach of the decision-makers to the value functions is attempted to explore. Hence, Table 4.1 will be referred to in the evaluations of the suppliers through the criteria set.

In the first statement in Table 4.1, the procurement managers are asked whether it is possible to evaluate the suppliers in general terms and act accordingly for every procurement decision. This statement was the one with the second lowest level of agreement among all the statements. The main objection here is that in an environment where the variety of products is so high, there is not a single best supplier which is preferred for every procurement decision. Instead, it is considered that working with more suppliers with different competence levels provides agility and a more robust supply chain. Hence, the values obtained from the suppliers should be analyzed in a more detailed way. The next two statements examine the resolution requirement of this assessment.

The responses to Statement 2 demonstrate that the evaluation procedure should be sensitive to mechanical parts because the average agreement score is 4.33 for this statement. In this manner, it appears that increasing resolution by including the product in the decision-making process is appropriate. After observing the responses to Statements 1 and 2, a follow-up question is directed to the participants whether it would be useful to reevaluate the suppliers in every purchasing request without using a predefined method. However, the procurement managers are not willing to increase the resolution to this level since the competencies of the suppliers do not significantly change in a short time. Therefore, they stated that the evaluation should take the specifications of the mechanical part into consideration, yet it should be done by a predefined method or score. Therefore, the typical procedure of AHP which is based on pairwise comparison of every alternative will not be applied in this problem due to practical issues.

The participants had the lowest level of agreement in Statement 3, which proposes to consider the competencies always higher the better. The important implication from the objections to this statement is that unnecessary competencies are not preferred, and there should be accordance between the competencies of the suppliers and the complexity of the mechanical part. In this manner, both agility will be gained by not occupying the capacity of highly competent suppliers and extra costs emerging from unused competencies will be avoided.

Upon the inputs taken from the procurement staff, a three-step procedure is determined. First, the data of the suppliers will be converted to a one-dimensional numerical scale. Methods of this transformation differ from criterion to criterion due to several reasons. One of them is the marginal value of the differences to the decision-maker. The second one is the number of dimensions of the original data. In other words, multiple attributes might have to be considered even for the evaluation of a sub-criterion. Last, the decision-makers are not equally precise about every criterion. In other words, a procurement staff might declare how much delay is to be tolerated in delivery, but it is more difficult to answer what should be the thresholds for the ideal cash ratio for this industry. Accordingly, clustering methods will be used to avoid determining predefined threshold values for these criteria. Instead, scoring will be determined based on the dispersion of the data. Specific motivations and methods for the criteria will be mentioned in the following subsections.

As a general explanation for determining the number of clusters, the motivation will be distinguishing the alternatives sufficiently, but not having too many clusters. In line with that for K-Means clustering, the number of clusters is recommended to be decided with Elbow Method (Thorndike, 1953). This method decides on the number of clusters based on the cost function given in Equation 4.1. Required notations of the equation are provided beforehand. The motivation behind this method is to increase the number of clusters until the decrease in J is sharply reduced between two consecutive numbers (Liu and Deng, 2021). The number of clusters might differ from criterion to criterion since the dispersions are different. Accordingly, the optimum number of clusters is determined for each bottom-level criterion separately by Elbow Method.

i : Cluster i

C_i : Centroid of Cluster i

k : Number of Clusters

x : Elements in Cluster i

$$J = \sum_{i=1}^{k} \sum_{x \in i} |x - C_i|^2 \tag{4.1}$$

In order to evaluate the clusters and assign scores to the suppliers in related clusters, the clusters will be ordered from worst to best. Then, a cluster will have a score of its cluster number divided by the number of clusters. In this way, the minimum score will be 1/N, and the maximum score will be 1, where N is the number of clusters for the criterion.

Even when the transformation mentioned above is completed, these outputs cannot be directly used since these sets of data only consider the attributes of the suppliers. To include the product attributes in the supplier selection process, the output of the first step will be used, and the suppliers' score of fitness to a mechanical part is found in Step 2. Afterward, a set of suppliers will be recommended based on the output from Step 2. The critical decision in Step 3 is to define a cutting point for the scores of the suppliers.

In §4.3.1, procedures for evaluating the alternatives concerning the criteria which have a quantitative scale is presented. In §4.3.2, an approach for nominal-scale criteria is developed.

4.3.1 Criteria with Quantitative Scale

Ratio and interval scales are quantitative as mentioned, but the difference is the existence of a true zero in the ratio scale. However, ratio scales might not be evaluated with linear relations in terms of the values of the alternatives. In other words, it can be said that a firm with 10 workers is twice the size of a firm with 5 workers, but the satisfaction from the second firm might not be half of the first one even if the workers are identical. Therefore, criteria with a ratio scale also require a procedure for evaluation rather than simply using the performance data itself.

While examining the evaluation process of the procurement staff, their sensitivity to the differences is questioned. The general approach was that small differences can be tolerated, but significant differences should affect the result. In other words, a decision-maker might be insensitive to the difference between two suppliers with Delivery scores of 73 and 74, but the divergence is desired between suitable suppliers and others. Ignoring the small differences is not a new finding in decision-making processes since there are already some MCDM methods that adopt this approach such as ELECTRE. Nonetheless, when this approach is combined with a desire for divergence between different alternatives, the purpose becomes grouping similar alternatives and forming these groups significantly different from each other. These motivations also support the use of clustering methods.

Machine Infrastructure: Evaluating the machine infrastructure of a firm is one of the most difficult parts of the supplier selection process since it is an MCDM problem itself. In other words, evaluating the firms will require multiple criteria such as the number of CNC machines and the number of axis of these machines. For instance, it can be safely said that a firm having two 5-axis CNC milling machines is more competent than a firm with only one 2-axis CNC Lathe Machine if the machines are equivalently good in their categories. Nevertheless, it is not that certain all the time due to the infinitely many possible machine parkour. Although the suppliers cannot be expressed as a single number with respect to this criterion, it will be evaluated with the number of machines, which is a set of numbers. Therefore, this criterion is considered to have a quantitative scale.

One obvious solution to this evaluation might be the pairwise comparison of all alternatives once and using these values for a time period, yet in such a situation that there are more than 400 suppliers, this would not be feasible. On the other hand, the powerful background of pairwise comparison might help to quantify the competencies of the firms. At this point, the variety of the machine parkours of the suppliers is investigated by asking the procurement staff how different can two suppliers have a different set of machines. The answer is that it is almost impossible to say any two suppliers have identical machine parkours, yet they are a set of machines that are used for the same purpose in the end. Hence, a unique machine parkour is observed very rarely, and the parkours of the suppliers are somewhat similar. Accordingly, procurement staff can evaluate the machine infrastructure of a firm if data is provided, even if they do not know the supplier before.

In this evaluation process, the AHP method will be applied in a distributed manner, with missing data. In this way, every procurement staff will compare a set of a reasonable number of alternatives. In order to prevent biases of the decision-makers, supplier names will not be presented. Instead, only the machine infrastructure of the firm should be shown for obtaining the comparisons. For each decision-maker, 6 alternatives will be compared. For the assignment, connectedness will be checked, and the suppliers will be reassigned if the connectedness is not assured. The features of CNC machines that need to be provided to the decision-makers are equipment category, brand, model, number of this machine, year, maximum spindle speed, number of axes, and axis limits.

A total comparison matrix will be formed as follows. The diagonals of the matrix will be equal to 1. For other elements, the geometric mean of the comparisons from different procurement staff will be obtained. If no one compared the pair of suppliers, 0 will be placed in the associated position. In the end, the geometric mean method (Crawford, 1987) will be used as a practical method to derive the priorities, by ignoring zeros in the rows.

Since this method is an application of AHP, all the suppliers will have a priority between 0 and 1, whose sum is 1 in the result. Therefore, the numerical values associated with the suppliers will also depend on the total number of suppliers. In order to deal with this situation, the score of the supplier with the highest score will be normalized to be equal to 1. Accordingly, ratios between the priorities of the firms will be preserved. For instance, in an environment where the lowest score is 0.05 and the highest one is 0.2, these scores will be normalized as 0.25 and 1. Since there is no provided data on the suppliers for machine infrastructure that is available, an illustrative example will not be provided in this study. A more detailed study on distributed AHP with missing data is conducted in §4.3.2, and explanations in this part are kept short to save space. For further information about incorporating the comparison matrices and assignment methodology, the evaluation procedure of the Communication criterion can also be seen.

Special Processes Competence: For this criterion, the procurement staff are asked what special processes are required for your mechanical parts most commonly. In this way, the scope of the evaluation is narrowed down to three special processes which are heat treatment, coating, and painting. As stated in §4.2, these special processes can be outsourced by the suppliers to the firms which are specialized in these operations; however, it is not preferred unless it is indispensable due to the complexity and quality issues. Then, their considerations about the evaluation procedure are received. The participants have reached a consensus that the existence of this ability is sufficient, so there is no need to score how well a supplier performs the coating process, for instance. Hence, every special process will have a weight in this evaluation, and the firms with these competencies will collect the associated weight as a score.

When the importance levels of these special processes are asked of the procurement staff, it is stated that they are almost equally important. However, running the heat treatment and coating operations is considered very slightly more important than the painting operation due to capacity issues. Afterward, it is proposed to assign the painting, coating, and heat treatment special processes the weights of 0.30, 0.35, and 0.35, respectively. These weights are considered acceptable by the participants. As an example, a supplier which is able to perform painting and heat treatment operations will get %65 (0.30 + 0.35) of the global weight of this sub-criterion, which is 0.65 * 0.036 = 0.0234.

Experience: The experience of a worker is often expressed in units of years. Although experience is critical in the manufacturing environment, the participants of the study believe that the acceleration of learning is higher at the beginning of the career. In other words, the difference between year 0 and year 2 is considered to be higher than the difference between years 20 and 22.

Since this study aims to evaluate the experience level of the workers at a supplier, this analysis will be conducted based on the average of the experience levels of the workers. To keep the process simple, experience levels are broken down into ranges. While so, value differences between two consecutive ranges are kept equal, but the difference between upper and lower bounds is incremented by 1. The scale of this criterion is defined as between 0 and 1, and associated scores with ranges are presented in Table 4.14, which is accepted by the participants.

Table 4.14: Evaluation Table of Experience

Average Experience (Years)	0-1	1-3	3-6	6-10	10-15	15+
Score	0	0.20	0.40	0.60	0.80	1

Number of Workers who can Use CAM: The workers who can use CAM Software are the required workforce for the machining processes with CNC Machines. Since there has to be at least one CAM operator for being a supplier of mechanical parts, suppliers with one worker who can use CAM software are considered compulsory and gets no credit. Similarly to the Experience criterion, the evaluation will be made with ranges, and the difference between upper and lower bounds is incremented by 1 between two consecutive ranges. The scale of this criterion is defined as between 0 and 1, and associated scores with ranges are presented in Table 4.15.

Table 4.15: Evaluation Table of CAM Operators

Number of CAM Operators	0-1	2-3	4-6	7-10	11-15	16+
Score	0	0.20	0.40	0.60	0.80	1

Annual Workforce Turnover: Building a dependable partnership with a supplier is important but not easy in such an environment where the complexity of both business processes and mechanical parts is high. Besides, these relationships are made through the workers in the suppliers. If the turnover ratio is above some threshold, these relationships are not solid since there will be always new contact people and production staff in the supplier.

For the organization, it is better if the suppliers can keep their workers within their firms. Hence, this criterion will be considered lower-the-better with an ideal value of 0. Considering the turnover ratio of the organization, an upper limit is determined as %30. Beyond this limit, the supplier will be evaluated as having an unacceptable turnover ratio, and get no credit. In between these two values, the score of the supplier with respect to turnover criterion is linearly decreased. Consequently, the score of the supplier for this criterion is calculated as in Equation 4.2.

$$Score = \max\left\{0, \frac{0.30 - TurnoverRatio}{0.30}\right\} \tag{4.2}$$

Number of Workers: From the manufacturing perspective, the workforce of the suppliers is categorized into four which are engineers, other university graduates, operators, and other workers. In this way, the evaluation is made with respect to four categories separately. In this criterion as well, a table is used for evaluating the suppliers. The ranges are equivalent to the evaluation of the CAM Operators, yet the values gained from the highest number are divided. In this way, if a supplier has the maximum number of workers in each category, the total score will be 1, but the workers from different categories are not substitutes.

For dividing a total score of 1 into four categories, engineers are considered to have more impact on the manufacturability of the items. Therefore, a maximum total score of 1 is divided into four as 0.3, 0.25, 0.25, and 0.20, respectively. The evaluation ranges and associated scores are presented in Table 4.16. A supplier will be evaluated with respect to four rows of the table, and the sum of the values obtained from each row will be the total score for this criterion.

Table 4.16: Evaluation Table of Number of Workers

Category	**0-1**	**2-3**	**4-6**	**7-10**	**11-15**	**16+**
Engineers	0	0.06	0.12	0.18	0.24	0.30
Operators	0	0.05	0.10	0.15	0.20	0.25
Other University Graduates	0	0.05	0.10	0.15	0.20	0.25
Other Workers	0	0.04	0.08	0.12	0.16	0.20

In line with the evaluation table, some supplier data are generated and scored as an illustrative example. The scores of these suppliers with respect to the criterion are calculated in Table 4.17. The motivation of this scoring method is to promote the suppliers with a higher workforce. On the other hand, if there is a need for trade-off, the firms with a more balanced workforce among categories are promoted. One example of this situation is the comparison of Suppliers 3 and 4 in Table 4.17, which have an equal number of workers but different distribution to the categories.

Historical Quality Performance: In the criteria related to historical performance, often there are multiple observations on the performance. In other words, the score of a supplier with respect to quality performance depends on multiple deliveries. Therefore, a set of scores from the deliveries should be combined for obtaining a single historical quality performance score. For this purpose, procurement staff is asked for their opinion

Table 4.17: Illustrative Example of Evaluation with respect to Number of Workers

	Number of Workers				
Suppliers	**Engineers**	**Operators**	**Other University Graduates**	**Other Workers**	**Total Score**
Supplier 1	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	0.00
Supplier 2	16 (0.30)	16 (0.25)	16 (0.25)	16 (0.25)	1.00
Supplier 3	7 (0.18)	7 (0.15)	25 (0.25)	4 (0.08)	0.66
Supplier 4	7 (0.18)	16 (0.25)	16 (0.25)	4 (0.08)	0.76
Supplier 5	11 (0.24)	16 (0.25)	16 (0.25)	4 (0.08)	0.82

on how to define the importance levels of different observations. At first, the total cost of the order and the amount of the delivery are considered appropriate weighting methods. However, they reached a consensus on every order is equally important since a quality problem for even one item with low cost might directly halt production in an assembly environment.

The quality performance of a product can be categorized into four which are no faults, minor faults, rework-needed faults, and unusable items. A no-fault case is the desired output which means there is no problem with the product. A minor fault means that there is no problem with the functionality, but the specifications are not completely met. For instance, if the color code of an item does not perfectly match its requirement, there is no problem with using this item but the supplier could not follow the requirement. Mechanical parts which require a rework operation have some problems and the organization does not prefer to use these items. However, the specifications could be met with rework operations. In the end, these items are delivered as no-fault items, yet the suppliers are expected not to deliver the parts which do not meet the specifications. Last, some mechanical parts cannot be used and rework operations are useless for these items. These parts are called unusable items, and this quality level is not acceptable at all.

For scoring the quality performance of an order, a no-fault case is the perfect output and it should get full credit. On the other hand, the delivery of unusable items is completely unacceptable and does not deserve credit. In this way, two endpoints of the scale are determined as 0 and 1. To place the other two cases on the scale, the opinions of the procurement staff are asked. As a result, the minor-fault case is considered very close to the no-fault case since it does not halt production, and it might not be realized by the supplier. However, a rework-needed case is much more problematic since it might halt production, cause a delay in the deliveries of the organization, and lead to extra costs due to additional operations. In light of these inputs, the scale in Table 4.18 is recommended to the procurement staff, and it is accepted.

Table 4.18: Evaluation Table of a Delivery in Terms of Historical Quality Performance

Product Quality	Unusable Items	Rework Needed	Minor Fault	No-Fault
Score	0	0.45	0.80	1

By using this table, each delivery is evaluated separately by a weighted average of the associated scores of the mechanical parts. To illustrate, if a delivery of 4 items is made to the organization in which each case is observed once, the score of this delivery is calculated as in Equation 4.3. Afterward, the arithmetic average of these single scores is considered the Historical Quality Performance of a supplier.

$$Score = \frac{1*0+1*0.45+1*0.80+1*1}{1+1+1+1} = 0.5625 \tag{4.3}$$

Quality Control Capabilities: As discussed in §4.2, a supplier with a successful history of quality performance might not be able to deliver more complex mechanical parts since the supplier might have been producing only parts with low complexity in the past. Moreover, if the tolerance levels are tight, besides the manufacturing process, even measuring the dimensions of the mechanical part is difficult. Although there are various tools for quality control, the procurement staff consider CMM to be the most important one for these mechanical parts manufactured by machining processes. Therefore, the CMM devices of the supplier will be evaluated in this criterion. For this purpose, the CMM models of different brands which are most commonly used and preferred by the organization are listed. Then, these models are categorized into four according to their perceived performances by a group of procurement staff.

For the measurement process of a mechanical part on CMM devices, a serious capacity problem is not foreseen. Therefore, it is predicted that the suppliers will be able to make appropriate assignments between the mechanical parts and CMM devices. In other words, if a supplier has more than one CMM device, it is expected to perform the measurement process of more complex parts with the one having more sensitivity. Therefore, the suppliers will get the maximum among the associated scores of the categories of the

CMM devices. Uncategorized CMM models are treated as ”others”, and they are evaluated in Category 0. Scores of the categories from 0-4 are presented in Table 4.19. As an illustrative example, six different suppliers are created and scored in Table 4.20.

Table 4.19: Evaluation Table of Quality Control Capabilities

CMM Categories	Category 0	Category 1	Category 2	Category 3	Category 4
Score	0.35	0.70	0.80	0.90	1.00

Table 4.20: Illustrative Example of Evaluation with Respect to Quality Control Capabilities

	Number of CMM Devices in Categories 0-4					
Suppliers	**0**	**1**	**2**	**3**	**4**	**Total Score**
Supplier 1	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.90)	1 (1.00)	1.00
Supplier 2	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (1.00)	1.00
Supplier 3	0 (0.00)	0 (0.00)	1 (0.80)	0 (0.00)	0 (0.00)	0.80
Supplier 4	0 (0.00)	0 (0.00)	1 (0.80)	0 (0.00)	3 (1.00)	1.00
Supplier 5	0 (0.00)	0 (0.00)	1 (0.80)	3 (0.90)	0 (0.00)	0.90
Supplier 6	1 (0.35)	1 (0.70)	0 (0.00)	0 (0.00)	0 (0.00)	0.70

Quality Certificates: The processes of the firms are considered to be complying with the standards in the subjects of the certificates if the firm is certified by an independent organization. In this way, these standards are the source of quality processes and quality outputs. Related quality certificates are determined as ISO 9001 and AS 9100, with equal importance levels. While ISO 9001 covers the standards of quality management independent of the industry, AS 9100 defines the design and manufacturing standards for the aviation and defense industry. Since the certification process considers various criteria about the business processes, only the existence of these certificates is evaluated. In line with that, a firm with both certificates will get a score of 1, a firm with one of these certificates will have a score of 0.5, and the firms with none of these certificates will not get any credit from this sub-criterion.

Historical Delivery Performance: The evaluation of this criterion has some similarities with the evaluation of Historical Quality Performance since there are multiple observations of the historical performance in both criteria. In this sense, this evaluation procedure will also combine multiple scores derived from different deliveries. Similarly, the weights of different deliveries will be equal since a delay in the delivery of even a small mechanical part in the bill of material can lead to a delay in the deliveries of the organization. Therefore, there is no delivery with less importance than others.

Focusing only on the delays in the deliveries would be an oversimplification since the delivery time is a two-way evaluation criterion. For instance, if a supplier demands to deliver an item one year earlier, it will not cause a halt in production and delay in the delivery, yet it will cause a big inventory cost. Besides, the organization will have to take all the risks of material handling of this item for one year. Hence, the deliveries from the suppliers are expected to be made on time, neither early nor late.

Due to the complexity of the business process, suppliers are welcome to make their deliveries 30 days before or 15 days later than the committed date. The limit of early delivery is wider since the inventory-related costs are considerably more acceptable than the risks of late delivery. Beyond these limits, suppliers will get credit within the 90-day deviation from the committed date, but beyond these limits, the delivery is considered unacceptable. Between these limits, the score is linearly decreased towards the endpoints. To conclude, if $x \in \{-\infty, \infty\}$ is the lateness of the delivery, the single delivery score is calculated as the piecewise function in Equation 4.4. Then, the arithmetic average of these scores will give the historical delivery performance score of the supplier.

$$Score = \left\{ \begin{array}{ll} 0, & \text{if } x < -90 \\ 1 - \frac{|x|-30}{60}, & \text{if } -90 \leq x < -30 \\ 1, & \text{if } -30 \leq x \leq 15 \\ 1 - \frac{x-15}{75}, & \text{if } 15 < x \leq 90 \\ 0, & \text{if } 90 < x \end{array} \right\} \tag{4.4}$$

Organization's Share in Supplier's Total Revenue: This criterion directly results in a ratio that lies on a scale between 0 and 1. If this ratio is equal to 1, it can be interpreted that all the revenue of the supplier is generated from the organization. Then, the supplier is expected to align its processes with the needs of the organization. On the other hand, if this ratio is close to 0, it can be understood that a very small portion of the revenue of the supplier is generated from the organization. In this case, the organization's requests may not be the priority of the supplier. Therefore, this criterion will be evaluated with its own ratio.

Cash Ratio: Evaluating the financial strength of a firm can be made with various metrics of which some are very similar with slight differences but some others are focusing on different aspects. These intense studies are the demonstration of the fact that evaluating the financial strength of a firm is not a simple and straightforward task. For strategic-level decisions, the organization explores all these metrics for the supplier. However, placing a procurement order to a supplier is a lower-level decision, so the focus is on whether the supplier will be able to afford the necessary costs for producing this item. Therefore, the evaluation procedure will be structured accordingly.

Although a too-high cash ratio might be interpreted as the firm cannot efficiently manage its cash, it still means that the firm has enough resources to pay its short-term debts. In line with that, this metric will be evaluated as a higher-the-better criterion. The next step is to determine the value association between the ratio and value of the suppliers gained from this criterion. Two widely accepted thresholds are 0.5 and 1. Generally, if this ratio is less than 0.5, it is considered risky, and a ratio greater than 1 is preferred. Nevertheless, there are two main risks to using these thresholds. The first one is having a low resolution for evaluating the suppliers, i.e., hundreds of firms are categorized into three, so there is a risk of unawareness of important differences. The second risk is that financial ratios vary between industries. Hence, general assumptions might not hold for this case. These concerns are valid for the other two sub-criteria of the Finance criterion. Thus, numerical data will be clustered first, and clusters will be evaluated to comply with these issues. Then, the suppliers will gain the score associated with their clusters.

Ability to Find Loan: Finding a loan is an ability of the supplier since it requires a clean financial history to ensure the trust of the firms which are willing to lend money. This ability is evaluated with respect to the Findeks Score, which is a risk-related metric in Turkey. While higher scores imply a low-risk level, firms with lower Findeks scores can be considered to have a higher financial risk. This common metric in Turkey has an important impact on finding a loan. The metric lies in a range between 1 and 1900. A general and accepted method for interpretation of this score is to evaluate the risk of the firms in five categories. The categories and associated Findeks Scores are presented in Table 4.21.

For this criterion, similar concerns to the ones in Cash Ratio exist. A firm with a

Table 4.21: Risk Categories and Findeks Scores

Category	High Risk	Medium Risk	Low Risk	Good	Very Good
Score	1-699	700-1099	1100-1499	1500-1699	1700-1900

Findeks score of 1699 might be considered less risky than another with a score of 1500. However, they will be evaluated with the same risk level with respect to the table above. Besides, the suppliers of the organization might be agglomerated into some subset of this range. Therefore, the suppliers will be clustered first, and their clusters will be evaluated in this sub-criterion as well.

Number of Customers: This attribute of the suppliers is expressed in numbers. However, a procedure should be determined for the relationship between the numbers and values gained from this sub-criterion. The customers are the sources of income for firms. Therefore, the evaluation will be made on a higher-the-better approach, but the ranges and associated values should be determined. However, this decision is not easy since the number of customers depends on the industry. For instance, an e-commerce firm might have thousands of customers and it is considered unsuccessful, yet one manufacturing firm might be considered very successful with ten customers. In line with that, similar to the other two sub-criterion of Finance, the suppliers are clustered first, and the clusters will be evaluated in this sub-criterion as well.

Total Recordable Incident Rate (TRIR): By TRIR metric, the number of incidents is normalized to an imaginary firm that works for 200,000 man hours. Although this metric is useful for evaluating firms of different sizes, the observation of this metric might highly depend on the industry. Hence, using the general thresholds for this metric for the manufacturing industry might be Despite the fact that desired observation for this sub-criterion is zero since every incident is critical, it is also important to separate relatively less successful suppliers from others. Therefore, the TRIR values of the firms will be collected, and they are clustered similarly to the sub-criteria of Finance.

Waste Management Policy: The consciousness of the firms regarding the management policies of waste is on different levels. While some firms have a well-planned procedure for handling waste, some have no procedure at all. Under these circumstances, it is certain that the firm with plans should be promoted, yet these firms are not performing waste management equally well either. Nevertheless, only the existence of a well-defined procedure is analyzed in this sub-criterion for two main reasons. The first one is that qualitative evaluation of different processes of different suppliers requires an objective and wise system, which is considerably laborious. The second one is that this sub-criterion is the one with the lowest global weight among the 18 lowest-level criteria. When these two reasons are considered together, evaluation of only the existence of this policy seems to be sufficient for the time being.

Management System Certificates on the Subject: In this sub-criterion, the existence of two desired standards of the International Organization for Standardization (ISO) is evaluated. The first one is ISO 14001, which defines the criteria of the environmental management system. The firms entitled to receive this certificate are evaluated as managing their business processes environment-friendly. The second standard is ISO 45001 which regards occupational health and safety subject. The firms having this certificate are considered to take necessary actions proactively. When the procurement staff are asked for their opinion, they stated that occupational health and safety studies can have a higher impact. The motivation behind this prioritization is that any negligence regarding occupational health and safety might lead to serious consequences in the short term. Then, the total value of 1 is divided into two as 0.4 and 0.6 respectively, as it is given in Table 4.22. The firms will gain the sum of scores of the certificates that they have.

Table 4.22: Certificates and Associated Scores

Certificate	ISO 14001	ISO 45001
Score	0.40	0.60

4.3.2 Criteria with Nominal Scale

Among 18 bottom-level criteria, only Communication cannot be quantified. Moreover, it is not possible to order the suppliers with respect to their Communication skills with respect to some metrics. Therefore, this evaluation cannot be handled by mathematical expressions. Accordingly, the only way is to receive the judgments of the decision-makers. At this point, another problem arises that there is a many-to-many relationship between the procurement staff and the suppliers. In other words, a procurement staff works with multiple suppliers, but not all of them. Moreover, the suppliers are in contact with multiple procurement staff from the organization. Hence, one procurement staff cannot evaluate all of the suppliers, and a supplier should be evaluated by multiple procurement staff. Last, the number of suppliers is too high to be evaluated with AHP due to the growth of the number of questions with respect to the number of elements to be compared. Consequently, applying AHP in a traditional way is not a solution.

If the discussion above is expressed within the AHP context more formally, the inability of the procurement staff to evaluate all of the suppliers causes missing data in the comparison matrix. The causes of missing data are similar to two of the reasons stated in Harker (1987), which are the decision-maker is not able to make the comparisons for some pairs, and the time to complete n(n-1)/2 comparisons. Besides, this evaluation process should be performed in a distributed manner since there are multiple judgments from multiple procurement staff regarding the communication success of the supplier on the organization's side. The evaluation of the suppliers for this criterion will be made by using pairwise comparisons.

For this evaluation, the solution of Harker (1987) and SEM (Oliva et al., 2017) will be applied. In addition, the geometric mean method will be applied as a computationally simpler alternative. For all weight derivation with missing data methods require the connectedness of the alternatives. This issue will be resolved by aggregating individual judgment matrices of the individuals. In other words, a classical AHP study should be conducted individually with all procurement staff to obtain pairwise comparisons of a set of suppliers that they are working with individually. By using the comparisons from these separately conducted AHP studies, a Sparse Group Comparison Matrix (SGCM) is obtained by aggregating individual judgments. Thanks to making assignments between the alternatives and decision-makers by considering connectedness, all suppliers will be connected through the comparisons of multiple decision-makers. In this manner, this will be an implementation of distributed decision-making.

Originally, a comparison matrix is two-dimensional in which both rows and columns represent a set of criteria or alternatives. In this problem, each individual will make their comparisons through a similar two-dimensional comparison matrix for a subset of the suppliers. However, the third dimension appears in this problem as individuals. Say i and j are the indices for alternatives and k is the index of individuals. The elements of SGCM will be computed as the geometric mean of the judgments of the individuals who made a comparison between associated alternatives (Oliva et al., 2018). The motivation for using geometric mean is that this is an aggregation on a judgment basis. Consequently, SGCM is also a two-dimensional comparison matrix, with possibly missing data. Other notations are provided below, and the expression of SGCM is given in Equation 4.5.

C : Set of Comparison Matrices of Individuals with Missing Data

N_{ij} : Set of Individuals who Made a Comparison between Alternatives i and j

$$SGCM(i,j) = \left\{ \begin{array}{ll} 0, & \text{if } |N_{ij}| = 0 \\ \left[\prod_{k \in N_{ij}} C_{kij}\right]^{\frac{1}{|N_{ij}|}}, & otherwise \end{array} \right\} \tag{4.5}$$

In the next subsection, the representativeness of these three methods will be tested. For this purpose, a set of complete comparison matrices will be used, and some of the elements will be treated as missing data. Then, these three methods will be applied to derive the priority vectors. In this real-life project, conducting such a large-scale AHP study is laborious. Therefore, it is preferred to test these methods with an available data set and take action accordingly. Details of the data set and testing methods will be explained in the next subsection.

4.3.3 Evaluation of the Methods for AHP with Missing Data

Ahmed and Kilic (2022) conducted two empirical AHP studies, which are comparing the density of dots in nine images and comparing the mass of nine different bottles. Accordingly, in both studies, the participants made 36 comparisons and formed a complete comparison matrix with a size of 9x9. In this study, the data set of visual comparison will be used. For the visual comparison study, 164 participants answered these 36 comparison questions. In this study, the visuals will be treated as the suppliers, and the participants will be considered as the procurement staff.

For this data set, some comparisons of the decision-makers will be treated as missing data, and only a few participants' data will be used to have sparsity in SGCM. In line with that, 3 alternatives are selected for each participant. Therefore, except for 3 of a total 36 comparisons will be treated as missing for every participant. In order to assure connectedness, the judgments of 6 sufficiently consistent decision-makers are taken as input. The first six participants of the study have a C.R. value less than 0.1, so these data will be used. Based on trials, random assignments of 3 alternatives to 6 decision-makers yield both connected and unconnected networks. Therefore, these numbers are considered enough to assure connectedness but still yield missing data, which is a desired observation.

The main motivation behind using this data set is that there are complete comparison matrices that are treated as incomplete. In other words, there is information about all preferences, and the suitability of the methods to incorporate partial information can be tested. This suitability term is a measure of how well these methods handle the matrices and approximate the results of complete comparison matrices consistently.

Saaty (1994) examines the compatibility notion as a metric of closeness. Two different vectors u and v, which store the measurements from the same scale are said to be compatible if the value of the compatibility metric is low. To compute the compatibility metric, consistent ratio matrices A and B which are obtained from the vectors are used. Ratio matrices are derived from the vectors by the following equations:

$$A_{ij} = u_i/u_j \tag{4.6}$$

$$B_{ij} = v_i/v_j \tag{4.7}$$

Compatibility between two matrices is defined as the sum of the elements in the Hadamard product of the first matrix and the transpose of the second matrix. The related expression is given in Equation 4.8.

$$c(u,v) = e^T(A \odot B)e = \sum_i \sum_j (A \odot B^T)_{ij} \tag{4.8}$$

In the evaluation of the communication performance of the suppliers, the purpose is to obtain a weight vector from the set of comparison matrices of individuals in which some data are missing. An appropriate method for obtaining this matrix incorporates the comparison data of individuals so that the obtained priorities are compatible with the individuals' comparisons. When the result of Equation 4.8 for two matrices is less, it can be inferred that these two matrices are more compatible. In addition, two perfectly compatible matrices yield a result of n^2.

In addition to the compatibility criterion, since there are multiple ways to make assignments between the suppliers and the procurement staff, it is preferred that the derived vectors are not too sensitive to the assignments. Otherwise, different assignments would result in very different priority vectors, which is an illustration of inconsistency. This inconsistency is not desired since the objective is to reach a robust result through less effort of decision-makers independent of who evaluated which supplier. In this sense, results for different assignments are desired to be close since conducting multiple iterations with different assignments in real-life applications is not easy. As the result of numerous studies with different assignments, it is expected to approximate an average value. However, little data will be gained in industrial applications due to the time requirements of conducting such a study. Therefore, a representative weight vector should also be close to the average of different iterations. Saaty (1994) provided an expression to measure the closeness of two vectors p and q, as in Equation 4.9.

$$\frac{1}{2}\left[\frac{p_1 p_2 \dots p_n}{q_1 q_2 \dots q_n} + \frac{q_1 q_2 \dots q_n}{p_1 p_2 \dots p_n}\right] \tag{4.9}$$

The methods adopted in this study which are Harker (1987), Oliva et al. (2017), and the Geometric Mean Procedure (1987) will be evaluated with respect to the compatibility and consistency criteria.

It is stated that there are 9x9 comparison matrices of individuals from the experiment. To test the success of the methods in terms of compatibility, results obtained from these three methods will be compared to the complete comparison case. In other words, if all comparisons are known, the comparison matrices are aggregated by element-wise geometric mean of the elements of different comparison matrices. The vector from this aggregated matrix is going to be compared to the one obtained by the three methods discussed above. Since different assignments will yield different results, these three methods will be applied for 100000 iterations to see the long-run behaviors. In each iteration, a different compatibility value is observed. As a result, the arithmetic average of the compatibility values over iterations will be reported. The related expression is given in Equation 4.10, and notations are given before.

W : Consistent Comparison Matrix Derived from the Weight Vector of Complete Comparison Matrices

t : Iterations

S_t : Consistent Comparison Matrix Derived from the Weight Vector of Sparse Comparison Matrices in Iteration t

$$c(w,s) = \frac{\sum_t e^T (W \odot S_t^T) e}{|T|} = \frac{\sum_t \sum_i \sum_j (W \odot S_t^T)_{ij}}{|T|} \tag{4.10}$$

One might argue that the performance of the methods should not be evaluated with the data which are treated as missing. Yet, the motivation of this whole process is to comply with the theoretical scenario that there is no missing data. Therefore, all the elements in complete comparison matrices will be included in the calculations.

For the consistency criteria, priority vectors obtained in iterations will be used. In the structure of Equation 4.9, the p vector will be considered as the arithmetic average of priority vectors obtained in iterations. The other vector will change in every iteration, which is the priority vector of the iteration. As a result, the arithmetic average of the consistency values over iterations will be reported.

The average consistency and compatibility values of these three methods are given in Table 4.23. Arithmetic averages of the weights obtained in iterations are presented in Table 4.24 for each of the methods. As a statistic, nearly 41 of 72 non-diagonal elements of the comparison matrix were missing in the iterations on average.

Table 4.23: Performance of the Adopted Methods

Methods	**Average Compatibility Value**	**Average Consistency Value**
Geometric Mean (Crawford, 1977)	92.64	1.37
Harker (1987)	87.49	1.50
SEM (Oliva et al., 2017)	87.33	1.51

Table 4.24: Arithmetic Mean of the Priorities from 100,000 Iterations

Alternatives	**Geometric Mean**	**Harker's Method**	**SEM**
1	0.273	0.316	0.316
2	0.205	0.220	0.220
3	0.152	0.150	0.150
4	0.107	0.098	0.098
5	0.083	0.071	0.071
6	0.061	0.050	0.050
7	0.050	0.040	0.040
8	0.041	0.032	0.032
9	0.029	0.022	0.022

According to the results in Tables 4.23 and 4.24, several interpretations can be put forward. First, Harker's Method and SEM yielded almost the same results. In fact, according to the first 3 decimals, generated priority vectors are exactly the same in the long run. In addition, the maximum priority is higher, and the minimum is lower in these two methods compared to the geometric mean method. Although these two performed better in terms of compatibility, the geometric mean method generated more consistent priority vectors. To conclude, consistency would not be an issue in the lack of compatibility, i.e., consistently generating less compatible vectors would not be desired. Therefore, Harker (1987) and SEM (Oliva et al., 2017) methods can be recommended according to this limited study. However, it should be stated that these observations should be validated with different data sets to reach a general insight.

Similarly to the evaluation of Machine Infrastructure, the maximum priority weight will be normalized to 1, and ratios between the alternatives will be preserved. For instance, if SEM is decided to be used, the credits of the suppliers gained from this criterion would be as in Table 4.25.

Table 4.25: Scores of the Suppliers for Communication Criterion

Alternatives	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Scores	1.000	0.696	0.475	0.310	0.225	0.158	0.127	0.101	0.070

4.3.4 Recommendation of the Suppliers

In this subsection, Step 2 and Step 3 of the solution methodology, which include finding a fitting score of the suppliers for a mechanical part and determining the subset to be proposed will be handled together.

As a common approach, MCDM methods handle the attributes either as a cost criterion or a benefit criterion. In this way, an alternative can be easily evaluated by either how low or high it is. In this manner, generally, there is no need for finding a fitting score for an alternative. Instead, normalized data or an associated score can be directly used.

In this study, the specifications of the mechanical part are also included in the supplier selection process. Moreover, there are two-way concerns about the competencies of the suppliers as discussed in §3. Therefore, the purpose is to select a supplier which is competent enough to produce the part but not too much to incur extra costs arising from unused competencies. Moreover, the organization needs available suppliers when a complex mechanical part is required.

Since the fitting scores are the last quantitative evaluation phase of the study, motivations concerning the criteria set should be reviewed. Starting with Technical Competencies, as previously stated the organization works with many suppliers with varying levels of competencies. It should be noted that these competencies are costly since they require a set of expensive equipment and qualified workers. For instance, prices of the CNC machines may vary significantly with respect to their quality, which leads to a higher cost. If a mechanical part has tight tolerances, these machines are required, and the organization is willing to accept additional costs. However, every part does not require this level of precision. In such a case, by benefiting from the diversity of the suppliers, simpler parts can be produced by the suppliers with fewer competencies. In this way, not only the cost structure becomes favorable to the organization but also working with more suppliers as a strategic decision can be attained.

In line with the above discussions, the Technical Competencies criterion is neither a cost criterion nor a benefit criterion. Instead, it is more of a target criterion. In other words, a set of highly competent suppliers are not the best alternative for all products. The best fit is determined by the conformity of the technical competencies of the firm to the complexity level of the product. Target-based evaluation is not novel in the literature. Jahan et al. (2012) used a target-based normalization in the TOPSIS method. In this way, the most favorable element is determined by closeness to the target value in the normalization stage of the method.

For target-based evaluation of the suppliers, a target should be specified, which reflects the complexity of the mechanical parts. To handle this issue, when a purchase requisition arrives the procurement staff will score the complexity out of N. In this way, the parts will be clustered into N groups. On the other side, the suppliers will be also clustered into N groups where the clusters are formed from 1 to N with increasing levels of competencies. This clustering will be made on the total score of the supplier from the Technical Competencies criterion. In this way, the firms and parts with the same cluster number are considered the best fit in terms of technical competencies. To illustrate, the suppliers from k^{th} cluster will have the highest fitting score from this criterion for the parts rated k out of N.

This clustering approach is applied to the suppliers separately for the other five main criteria as well. Cluster numbers are given in increasing order of the scores for these criteria as well. As a difference, the other five criteria are all benefit criteria because of their evaluation methods. Therefore, the fitting scores will be made with a higher-the-better approach. From the target point of view, the target value is determined as the number of clusters. The motivation here is that there is no desired level of delivery or communication. In other words, if two suppliers have identical equipment and workforce, the one with higher communication skills is preferred over the other for all of the mechanical parts as long as there is enough capacity. In this sense, the suppliers from N^{th} cluster will have the highest fitting score for all procurement decisions.

Assigning a fitting score to the best-fit suppliers is straightforward since they are evaluated as the best. In this manner, they will get full credit from related criteria. However, an appropriate function is required to evaluate others since there are often trade-offs between the alternatives in supplier selection. In other words, since there are multiple criteria, a supplier might appear to be the best alternative concerning the linear additive utility model even if it is not the best fit for any of the criteria.

Rezaei (2018) defines 10 different piecewise linear value functions with example cases for which each of these functions can be used. Among these 10 functions, only two are non-monotonic, which are V-Shape and Inverted V-Shape value functions. The Inverted V-Shape function is appropriate for the Technical Competencies criterion since this function is linearly increasing up to some level with a value of 1, and then linearly decreasing. The level with a value of 1 is specified as the target value, i.e., the cluster number of the mechanical part. As a variation, the deviations from the target will be penalized more in the levels below the target. The motivation here is that the cost of unused competencies is relatively more acceptable than the risks of choosing an incompetent supplier.

For the other five criteria, there are four different increasing piecewise value functions in the same study. Two of them are Level-Increase and Increase-Level in which the value remains constant within the range of level and is linearly increasing in the other range. These functions might be appropriate in the case of the existence of threshold values beyond which the decision-maker is not interested in the changes. The name of another one is Increasing, in which the value function increases linearly. Last, the Increasing Stepwise function is similar to the Increasing, but the value changes occur at certain levels. For this study, the Increasing value function will be preferred, yet since the cluster numbers are integers, it will be kind of a stepwise function. The motivation here is that adding thresholds as in Level functions might be useful in some cases but it is also important to distinguish the fitting scores of the suppliers. To illustrate, in the case of using a Level-Increase value function, if the endpoint of the Level is determined as 5, it is not possible to distinguish two suppliers whose cluster numbers are 1 and 5. A decision-maker might not concern if there are many suppliers with high fitting scores already; however, for some complex parts, there might not be a big number of suppliers.

Accordingly, the value function for the fitting scores concerning Technical Competencies is defined in Equation 4.11, and the value functions for the other five main criteria are defined in Equation 4.12. Notation that will be used in these equations will be provided beforehand.

I : Set of Suppliers

J : Set of Main Criteria

S_{ij} : Cluster Number of Supplier i for Criterion j

C_p : Cluster Number of the Mechanical Part p with respect to Its Complexity

F_{ij} : Fitting Score of Supplier i for Criterion j

F_i : Total Fitting Score of Supplier i

$$F_{i1} = \left\{ \begin{array}{ll} max\{0, N - 2 * |C_p - S_{i1}|\}, & \text{if } S_{i1} < C_p \\ N, & \text{if } S_{i1} = C_p \\ N - |S_{i1} - C_p|, & \text{if } S_{i1} > C_p \end{array} \right\} \tag{4.11}$$

$$F_{ij} = S_{ij} \tag{4.12}$$

As can be seen above, the minimum score is bounded by 0 for the Technical Competence criterion. One other interpretation concerning the fitting scores for Technical Competence is that the minimum scores for the parts with a medium level of complexity will be higher. To illustrate, when the suppliers and parts are clustered into 10 groups, clusters 1-5 will get no credit for a mechanical part belonging to cluster 10. On the other hand, every supplier will have a positive fitting score for a part from cluster 3, and the minimum score will be 3. This is a desired output since extremely simple and complex mechanical parts should be allocated to certain suppliers.

Afterward, a total fitting score of a supplier is calculated as the weighted sum of fitting scores from the criteria. Weights of the main criteria which are presented in Table 4.13 will be used in this summation. The mathematical expression for calculating the total fitting score is presented in Equation 4.13.

$$F_i = \sum_j w_j * F_{ij} \tag{4.13}$$

Calculation of total fitting scores brings the subject into the last step of the supplier selection process, which is how high the total fitting score will be sufficient to be on the shortlist for the recommendation. In other words, every supplier will have a total fitting score, but below a certain threshold should not be included. There might be several straightforward approaches such as recommending the first k of the list or the first p percent of the list. The risk here is that the distribution of the scores cannot be predicted. Therefore, the cutting point might exclude a supplier whose fitting score is very slightly less than the last supplier on the shortlist. One other simple approach can be to determine a threshold for fitting scores such as x out of N. In this case, the suppliers with a total fitting score of x or higher will be on the shortlist. Nonetheless, there are certain risks such as recommending less than the requirement or including too many suppliers on the list because of the uncertain distribution of the fitting scores.

To the problem of uncertain distribution of the scores, a clustering method is already proposed. A similar approach will be recommended here. As the first step, the total fitting scores of the suppliers will be clustered into N, and the optimum value of N will be decided with the Elbow Method. Afterward, the suppliers from the cluster whose centroid has the biggest total fitting score will be recommended first. If there are enough suppliers, the recommendation will stop. If not, suppliers from the next cluster will be recommended. This adding process will continue until a sufficient number of suppliers are included in the shortlist. In this way, cutting points will be lowered until a significant difference in total fitting scores is observed.

Chapter 5

CONCLUSION

In this study, a preceding stage of the evaluation of the quotes is studied. This problem has several difficulties such as some critical information cannot be known before asking the supplier. However, asking a supplier to send a quote is also costly. In this sense, the objective of this study is to recommend a minimum sufficient number of suppliers. For this purpose, both the current assets and competencies of the firm and historical performance values need to be used.

The product group whose supplier selection is studied is the mechanical parts that are produced with machining processes. Machining processes are complex since even the manufacturability term is related to various technical competence factors. Besides the competencies, it is also important that how the supplier run their business. Therefore, the supplier selection problem is handled as an MCDM problem in which various criteria are considered. Since these suppliers will produce a mechanical part designed by the organization, criteria of technical competencies occupy an important place. Nevertheless, quality and delivery criteria still appear as the most prioritized criteria even for this problem.

Although the global weight of the Environment and Safety criterion is determined as the lowest, this criterion is gaining more importance in the literature. Hence, in the future, this criterion is predicted to be divided and elaborated into a higher number of main and lower-level criteria. Moreover, the evaluation procedures can be more detailed as the impact of these factors is increased in terms of global weights. In this study, these weights are evaluated to be relatively low because the objective is not to make a strategic-level decision. Nonetheless, as the consciousness about these topics is enhanced, it is predicted that these factors will have reflections of the strategic-level decisions on more tactical and strategic-level decisions. Furthermore, new regulations might also force organizations to adjust their actions accordingly. Thus, this criterion can be the source for further studies.

Even for a single organization's procurement decisions, it is very difficult to build a system that covers all the possible outcomes. Even if it is done, then it would cause a vast complexity and maybe infeasibility. In order to relax the problem, there are some underlying assumptions in this study as well. For instance, the suppliers are evaluated in terms of historical delivery performance with respect to their deviations from the committed date. Nonetheless, this evaluation ignores the possibility of partial delivery. Adding this resolution to the problem requires a weighting method for different deliveries within an order as well. Moreover, some delays in the delivery might be caused by the organization such as approval processes and design changes. These discussions can be valid for every criterion in the hierarchy. Therefore, the evaluation procedure of every single sub-criterion is an opportunity for further studies.

Inputs from the procurement staff enlightened how the suppliers should be evaluated if pairwise comparisons are not used. When the staff can precisely define what they need and what they can compromise, threshold values are defined, and the suppliers are evaluated accordingly. On the other hand, the staff can not be precise in every criterion. For instance, a procurement engineer who purchases mechanical parts might not clearly say what is the maximum acceptable TRIR value. For such cases, the K-Means Clustering algorithm is proposed and by this method, the scores are determined according to the distribution of the data itself.

Besides the criteria, used MCDM method is also crucial since it determines the procedure of numerical operations. Since the supplier selection problem is a famous application field of MCDM, numerous methods are proposed and applied in the literature. In the end, using AHP with slight variations is preferred. In fact, the ease of changing some parts is one of the biggest advantages of AHP. Some other several advantages can be listed as usability in hybrid solutions, having widespread applications, and diversity in extensions.

Since the nature of the supplier selection problem is very diverse, traditional methods might be used with their variations. In this study, various extensions of AHP are used such as group decision-making and decision-making with missing data. For observing the performances of these extensions, an available data set is used instead of conducting an experiment. The motivation here is to compare the results from matrices with missing elements and complete comparison matrices. Since more than 400 suppliers cannot be evaluated by the traditional AHP approach, the preferred data set is convenient for this purpose. On the other hand, there are also some drawbacks to using this data set in a distributed decision-making problem. As an advantage, the available data set is from the AHP context, but there was an ordinal relation among the alternatives. In other words, if Alternative 1 includes more dots than Alternative 2, this relation will naturally hold for almost every decision-maker. Thus, consistency and compatibility levels might be higher than in the real world of AHP with missing data. Nevertheless, in the supplier selection context, the ordinal rankings of the suppliers might differ from person to person. According to the results of this study, Harker's Method (1987) and SEM (Oliva et al., 2017) performed better in terms of compatibility. On the other hand, the Geometric Mean Procedure (Crawford, 1987) generated more consistent priority vectors.

After the suppliers are scored with respect to each criterion, a need for a fitting score appeared in this study. The reason for this need is that one of the main criteria is a target criterion. In fact, there might be other reasons as well if the value functions can get more complex. Nonetheless, one of the most important objectives of this study is to build a simple system. Accordingly, two linear piece-wise value functions are determined and used in this study. Upon calculating the fitting scores of the suppliers, an approach is proposed to determine the suppliers that are asked for a quote. K-Means clustering is used in this approach as well.

Besides the potential further studies mentioned till now, some others can be listed as the following. The first one is finding criteria that are not evaluated based on the declaration of the suppliers. For instance, historical quality performance can be measured by the information within the organization. However, the data of the number of customers of the supplier is a statement of the supplier, which might be misleading. The second is conducting an experiment for distributed AHP with missing data. Moreover, alternative evaluation metrics for the success of the methods can be studied. Third, the preferences of the decision-makers can be explored by deduction instead of induction. In other words, the preferences of the procurement staff can be examined by artificial intelligence methods and past procurement order data. Fourth, the relation between the specifications of the mechanical parts and the competencies of the suppliers can be further investigated. In this study, they are matched based on their clusters, but every mechanical part is different.

Therefore, a more detailed study on this requirement might enable the decision-maker to eliminate the risk of working with incompetent suppliers, but this study would be on the machining topic. Last, Step 2 of this supplier selection study, which is computing the fitting scores can be developed based on the experience of the procurement staff.

To sum up, it is very difficult to close a study on the supplier selection topic. Even for this single study, there were many other research topics that can be included. On the other hand, there were some parts of this study that could have been handled within the scope of a different study. Therefore, it can be said that both ways of narrowing down or elaborating the scope are appropriate. In addition, it can be said that the supplier selection problem will continue to attract attention since new solutions to old problems will point out new problems. As a concluding remark, practical ease of implementation should always be considered in supplier selection studies because these systems seem likely to keep involving humans for a considerable amount of time.

BIBLIOGRAPHY

Aczél, J., & Saaty, T. L. (1983). Procedures for synthesizing ratio judgements. Journal of Mathematical Psychology, 27(1), 93–102. https://doi.org/10.1016/0022-2496(83)90028-7

Ahmed, F., & Kilic, K. (2022). A Basic Algorithm for Generating Individualized Numerical Scale (BAGINS). arXiv preprint arXiv:2211.08740.

Arrow, K. J. (1963). Social Choice and individual values. Yale University Press.

Barbarosoglu, G., & Yazgac, T. (1997). An application of the analytic hierarchy process to the supplier selection problem. Production and Inventory Management Journal, 38(1), 14-21. https://www.proquest.com/scholarly-journals/application-analytic-hierarchy-process-supplier/docview/199928362/se-2

Beil, D. R. (2011). Supplier selection. Wiley Encyclopedia of Operations Research and Management Science. https://doi.org/10.1002/9780470400531.eorms0852

Black, D. (1948). On the rationale of group decision-making. Journal of Political Economy, 56(1), 23–34. https://doi.org/10.1086/256633

Chai, J., & Ngai, E. W. T. (2020). Decision-making techniques in supplier selection: Recent accomplishments and what lies ahead. Expert Systems with Applications, 140, 112903. https://doi.org/10.1016/j.eswa.2019.112903

Chai, J., Liu, J. N. K., & Ngai, E. W. T. (2013). Application of decision-making techniques in supplier selection: A systematic review of literature. Expert Systems with Applications, 40(10), 3872–3885. https://doi.org/10.1016/j.eswa.2012.12.040

Chan, F. T., & Chan, H. K. (2004). Development of the supplier selection model—a case study in the Advanced Technology Industry. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 218(12), 1807–1824. https://doi.org/10.1177/095440540421801213

Chandran, B., Golden, B., & Wasil, E. (2005). Linear programming models for estimating weights in the analytic hierarchy process. Computers & Operations Research, 32(9), 2235–2254. https://doi.org/10.1016/j.cor.2004.02.010

Crawford, G. B. (1987). The geometric mean procedure for estimating the scale of a judgement matrix. Mathematical Modelling, 9(3–5), 327–334. https://doi.org/10.1016/0270-0255(87)90489-1

de Boer, L., Labro, E., & Morlacchi, P. (2001). A review of methods supporting supplier selection. European Journal of Purchasing & Supply Management, 7(2), 75–89. https://doi.org/10.1016/s0969-7012(00)00028-9

Ellram, L. M. (1995). Total cost of ownership. International Journal of Physical Distribution & Logistics Management, 25(8), 4–23. https://doi.org/10.1108/09600039510099928

Escobar, M. T., & Moreno-jiménez, J. M. (2006). Aggregation of individual preference structures in AHP-Group Decision making. Group Decision and Negotiation, 16(4), 287–301. https://doi.org/10.1007/s10726-006-9050-x

Fazlollahtabar, H., Mahdavi, I., Ashoori, M. T., Kaviani, S., & Mahdavi-Amiri, N. (2011). A multi-objective decision-making process of supplier selection and order allocation for multi-period scheduling in an electronic market. The International Journal of Advanced Manufacturing Technology, 52(9–12), 1039–1052. https://doi.org/10.1007/s00170-010-2800-6

Forman, E., & Peniwati, K. (1998). Aggregating individual judgments and priorities with the analytic hierarchy process. European Journal of Operational Research, 108(1), 165–169. https://doi.org/10.1016/s0377-2217(97)00244-0

Gencer, C., & Gürpinar, D. (2007). Analytic Network Process in supplier selection: A case study in an electronic firm. Applied Mathematical Modelling, 31(11), 2475–2486. https://doi.org/10.1016/j.apm.2006.10.002

Ghodsypour, S. H., & O'Brien, C. (1998). A decision support system for supplier selection using an integrated analytic hierarchy process and linear programming. International Journal of Production Economics, 56–57, 199–212. https://doi.org/10.1016/s0925-5273(97)00009-1

Harker, P. T., & Vargas, L. G. (1987). The theory of ratio scale estimation: Saaty's analytic hierarchy process. Management Science, 33(11), 1383–1403. https://doi.org/10.1287/mnsc.33.11.1383

Harker, P.T. (1987). Alternative modes of questioning in the analytic hierarchy process. Mathematical Modelling, 9(3–5), 353–360. https://doi.org/10.1016/0270-0255(87)90492-1

Ho, W., & Ma, X. (2018). The state-of-the-art integrations and applications of the analytic hierarchy process. European Journal of Operational Research, 267(2), 399–414. https://doi.org/10.1016/j.ejor.2017.09.007

Huang, S. H., & Keskar, H. (2007). Comprehensive and configurable metrics for supplier selection. International Journal of Production Economics, 105(2), 510–523. https://doi.org/10.1016/j.ijpe.2006.04.020

Ishizaka, A. (2012). Clusters and pivots for evaluating a large number of alternatives in AHP. Pesquisa Operacional, 32(1), 87–102. https://doi.org/10.1590/s0101-74382012005000002

Ishizaka, A., & Labib, A. (2011). Review of the main developments in the analytic hierarchy process. Expert Systems with Applications. https://doi.org/10.1016/j.eswa.2011.04.143

Jahan, A., Bahraminasab, M., & Edwards, K. L. (2012). A target-based normalization technique for materials selection. Materials & Design, 35, 647–654. https://doi.org/10.1016/j.matdes.2011.09.005

Kannan, V. R., & Tan, K. C. (2002). Supplier selection and assessment: Their impact on business performance. The Journal of Supply Chain Management, 38(4), 11–21. https://doi.org/10.1111/j.1745-493x.2002.tb00139.x

Karande, P., & Chakraborty, S. (2013). Using macbeth method for supplier selection in manufacturing environment. International Journal of Industrial Engineering Computations, 4(2), 259–272. https://doi.org/10.5267/j.ijiec.2013.01.002

Kenyon, G. N., Meixell, M. J., & Westfall, P. H. (2016). Production Outsourcing and Operational Performance: An empirical study using secondary data. International Journal of Production Economics, 171, 336–349. https://doi.org/10.1016/j.ijpe.2015.09.017

Lai, V. S., Wong, B. K., & Cheung, W. (2002). Group decision making in a multiple criteria environment: A case using the AHP in software selection. European Journal of Operational Research, 137(1), 134–144. https://doi.org/10.1016/s0377-2217(01)00084-4

Liu, F., & Deng, Y. (2021). Determine the number of unknown targets in open world based on Elbow Method. IEEE Transactions on Fuzzy Systems, 29(5), 986–995. https://doi.org/10.1109/tfuzz.2020.2966182

Liu, J., Ding, F., & Lall, V. (2000). Using data envelopment analysis to compare suppliers for supplier selection and Performance Improvement. Supply Chain Management:

An International Journal, 5(3), 143–150.
https://doi.org/10.1108/13598540010338893
Marcarelli, G., & Squillante, M. (2019). A group-AHP-based approach for selecting the best public tender. Soft Computing, 24(18), 13717–13724.
https://doi.org/10.1007/s00500-019-04479-1
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63(2), 81–97.
https://doi.org/10.1037/h0043158
Min, H. (1994). International Supplier selection: International Journal of Physical Distribution & Logistics Management, 24(5), 24–33.
https://doi.org/10.1108/09600039410064008
Nydick, R. L., & Hill, R. P. (1992). Using the analytic hierarchy process to structure the supplier selection procedure. International Journal of Purchasing and Materials Management, 28(2), 31–36. https://doi.org/10.1111/j.1745-493x.1992.tb00561.x
Oliva, G., Setola, R., & Scala, A. (2017). Sparse and distributed analytic hierarchy process. Automatica, 85, 211–220. https://doi.org/10.1016/j.automatica.2017.07.051
Oliva, G., Setola, R., Scala, A., & Dell'Olmo, P. (2018). Sparse Analytic Hierarchy Process: An Experimental Analysis. Soft Computing, 23(9), 2887–2898.
https://doi.org/10.1007/s00500-018-3401-9
Prasad, K. G. D., Prasad, M. V., Rao, S. V. V. B., & Patro, C. S. (2016). Supplier selection through AHP-Vikor Integrated Methodology. International Journal of Industrial Engineering, 3(5), 1–6. https://doi.org/10.14445/23499362/ijie-v3i5p101
Rezaei, J. (2015). Best-worst multi-criteria decision-making method. Omega, 53, 49–57.
https://doi.org/10.1016/j.omega.2014.11.009
Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. Omega, 64, 126–130.
https://doi.org/10.1016/j.omega.2015.12.001
Rezaei, J. (2018). Piecewise linear value functions for multi-criteria decision-making. Expert Systems with Applications, 98, 43–56.
https://doi.org/10.1016/j.eswa.2018.01.004
Rezaei, J., Nispeling, T., Sarkis, J., & Tavasszy, L. (2016). A supplier selection life cycle approach integrating traditional and environmental criteria using the best worst method. Journal of Cleaner Production, 135, 577–588.
https://doi.org/10.1016/j.jclepro.2016.06.125
Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology, 15(3), 234–281. https://doi.org/10.1016/0022-2496(77)90033-5
Saaty, T. L. (1980). The Analytic Hierarchy Process: Planning, priority setting, resource allocation. McGraw-Hill.
Saaty, T. L. (1990). How to make a decision: The Analytic Hierarchy Process. European Journal of Operational Research, 48(1), 9–26. https://doi.org/10.1016/0377-2217(90)90057-i
Saaty, T. L. (1994). A ratio scale metric and the compatibility of ratio scales: The possibility of Arrow's impossibility theorem. Applied Mathematics Letters, 7(6), 51–57.
https://doi.org/10.1016/0893-9659(94)90093-0
Saaty, T. L. (2001). Deriving the AHP 1-9 scale form First principles. Proceedings of the International Symposium on the Analytic Hierarchy Process.
https://doi.org/10.13033/isahp.y2001.030

Saaty, T. L., & Ozdemir, M. S. (2003). Why the magic number seven plus or minus two. Mathematical and Computer Modelling, 38(3–4), 233–244. https://doi.org/10.1016/s0895-7177(03)90083-5

Safarzadeh, S., Khansefid, S., & Rasti-Barzoki, M. (2018). A group multi-criteria decision-making based on best-worst method. Computers & Industrial Engineering, 126, 111–121. https://doi.org/10.1016/j.cie.2018.09.011

Sanayei, A., Farid Mousavi, S., Abdi, M. R., & Mohaghar, A. (2008). An integrated group decision-making process for supplier selection and order allocation using multi-attribute utility theory and Linear Programming. Journal of the Franklin Institute, 345(7), 731–747. https://doi.org/10.1016/j.jfranklin.2008.03.005

Sarkis, J., & Talluri, S. (2002). A model for strategic supplier selection. The Journal of Supply Chain Management, 38(1), 18–28. https://doi.org/10.1111/j.1745-493x.2002.tb00117.x

Stevens, S. S. (1946). On the theory of scales of measurement. Science, 103(2684), 677–680. https://doi.org/10.1126/science.103.2684.677

Taherdoost, H., & Brard, A. (2019). Analyzing the process of supplier selection criteria and methods. Procedia Manufacturing, 32, 1024–1034. https://doi.org/10.1016/j.promfg.2019.02.317

Thorndike, R. L. (1953). Who belongs in the family? Psychometrika, 18(4), 267–276. https://doi.org/10.1007/bf02289263

Verma, R., & Pullman, M. E. (1998). An analysis of the supplier selection process. Omega, 26(6), 739–750. https://doi.org/10.1016/s0305-0483(98)00023-1

Weber, C. A. (1996). A data envelopment analysis approach to measuring vendor performance. Supply Chain Management: An International Journal, 1(1), 28–39. https://doi.org/10.1108/13598549610155242

Yue, Z. (2011). A method for group decision-making based on determining weights of decision makers using Topsis. Applied Mathematical Modelling, 35(4), 1926–1936. https://doi.org/10.1016/j.apm.2010.11.001