THE CONTRIBUTION OF WORK STRESS TO FIRMS' EFFICIENCY: DATA ENVELOPMENT ANALYSIS BASED PSYCHOMETRIC CASE STUDY

ONUR ARIN ÖZTÜRK

Master's Thesis

Graduate School
Izmir University of Economics
2022

THE CONTRIBUTION OF WORK STRESS TO FIRMS' EFFICIENCY: DATA ENVELOPMENT ANALYSIS BASED PSYCHOMETRIC CASE STUDY

ONUR ARIN ÖZTÜRK

A Thesis Submitted to

The Graduate School of Izmir University of Economics

Master Program in Industrial Engineering

ABSTRACT

THE CONTRIBUTION OF STRESS TO FIRMS' EFFICIENCY:
DATA ENVELOPEMENT ANALYSIS BASED PSYCHOMETRIC CASE STUDY

Öztürk, Onur Arın

Master Program in Industrial Engineering

Advisor: Prof. Dr. Selin Özpeynirci

January, 2022

Efficient and optimal production is the main objective for firms disregarding their industry. Efficiency and productivity are variables which have lots of dynamics. As every dynamic, which requires human labor, efficiency and productivity are also affected by employees. According to psychologists, stress is the fundamental issue that affects work efficiency. Many studies indicate that stress has physical effects as much as mental effects on people. The aim of this study is to ascertain indirect effects of stress to production efficiency and productivity. With the aid of Data Envelopment Analysis and Item Response Theory, two different production plants located in Izmir, Turkey have been examined via Likert-type questionnaire specifically designed to measure work stress (WSQ) and via detailed efficiency analysis. Results indicate that stress affects not only employees but also firms' overall efficiency as well.

Keywords: Efficiency, Worker's Stress, Item Response Theory, Data Envelopment Analysis, Frontier Analysis

ÖZET

FİRMA VERİMLİLİĞİNE STRESİN ETKİLERİ:
VERİ ZARFLAMA ANALİZİ TEMELLİ PSİKOMETRİK VAKA ÇALIŞMASI

Öztürk, Onur Arın

Endüstri Mühendisliği Yüksek Lisans Programı

Tez Danışmanı: Prof. Dr. Selin Özpeynirci

Ocak, 2022

Verimli ve optimal üretim sektör gözetmeksizin bütün firmaların ortak hedefidir. Verimlilik ve üretkenlik dinamiği çok fazla olan değişkenlerdir. İnsanın olduğu her alan gibi verimlilik ve üretkenlik de dolaylı yoldan çalışan performansına bağlıdır. İş verimliliğini etkileyebilecek insan kaynaklı en temel unsur stres olarak görülmektedir. Araştırmalar, stresin insan üzerinde zihinsel olduğu kadar fiziksel etkileri olduğunu da göstermektedir. Bu çalışmanın amacı stresin insanlar üzerindeki dolaylı etkilerini incelemek ve üretim sektöründe çalışanların iş kaynaklı yaşadıkları stresin verimlilik ve üretkenliğe dolaylı etkilerini gözlemlemektir. Veri Zarflama Analizi ve Madde Tepki Kuramı ile Türkiye'de bulunan iki ayrı üretim fabrikasının çalışanları üzerinde yapılan anket uygulaması ve verimlilik analizleri stresin sadece çalışanların değil aynı zamanda firmanın verimliliğine de zarar verdiğini göstermiştir.

Anahtar Kelimeler: Verimlilik, İş Stresi, Madde Tepki Kuramı, Veri Zarflama Analizi.

To my sister...
Because we are never one without the other.

ACKNOWLEDGEMENTS

First of all, I would like to appreciate my advisor Prof. Dr. Selin Özpeynirci. Her support and wisdom always inspire and impress me and I hope I can be as successful as she is that she would be proud of me in the future.

I am so grateful for two firms and their employees for willingly participating to my thesis.

Also, special thanks to Assoc. Prof. Dr. Fredrik Ødegaard for aiding my thesis project with his knowledge and Prof. Dr. Kristina Holmgren for allowing me to use WSQ for my thesis. Their work inspired my project a lot and I feel very blessed to have a chance to meet with them.

Finally, I would like to thank to my family and friends. Along the way, I always felt their encouragement. Without them this chapter of my life would have been harder than it was.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Stress is a reaction that living creatures gives for the requirement of any adjustments or response. It can occur in many forms like mental, physical or emotional. Stress is a part of our lives in many areas, as it can be observed as positive and negative (Matthews, 2000). Stress can drive individuals to meet a deadline or maybe prevent themselves from a dangerous state, which can be a positive effect. However, stress most likely emerges as a negative phenomenon especially at work. Work related stress is a rising subject of interest since the mental-illness frequency is drastically increases among the employers as well as employees (Black, 2008). It can undermine the success of both individuals and organizations, some problems caused by stress are increased absenteeism and turnover, reduced quantity and quality of work, reduced job satisfaction and morale, problems of recruitment, poor communication and increased conflict (Michie, 2002). These outcomes have significant impact on firms' profitability and productivity. According to State of Global Workplace, 85% of employees disengaged or not focused on the work they are performing can cause almost $7 trillion to the company (Harter, 2017). Additionally, American Institute of Stress reports that roughly 12% of people have called sick due to job stress consequently, Stress related sickness causes business to lose almost $300 billion which has adverse impact on some firms and even could lead to bankruptcy (Business News Daily, 2020). Due to its cruciality, numerous firms initiated to launch their own policy combining their firm standards and the policies that are created by the institutes like World Health Organization and precisely workplace health-oriented organization like European Network For Workplace Health Promotion. It is another proof that work related stress not only causes individual inefficiency but also the work efficiency and productivity are affected affluently. This indicates that stress-free work environment provides less input, more output of better quality, and better performance overall (Blumenfeld and Inman, 2009). In that sense, it is crucial to answer, how is the productivity and production efficiency is affected by the worker's stress levels? What performance measures must be considered, regarding the existence of stress, to measure productivity and production efficiency? In light of the answers for the research questions, firms may consider measuring the performance in terms of organization productivity with the enhancement of potential gain that will be provided by labor quality attribute: Stress.

To analyze the impact of stress, it is offered to consider the measurement process to be separated. The actual production process that an organization uses to measure the productivity and efficiency and additionally the labor quality process. The actual production process expresses the operational production of goods and services, production planning process, decisions about production technology etc. The labor quality process indicates the supportive activities that aim to improve the organizational efficiency and productivity through promotion of labor quality attributes (Ødegaard and Roos, 2014). This case specifically refers to worker's stress levels, hence the labor quality process will be called stress affected process in this study. Although, worker's stress levels may not have direct impact on production processes, it is believed that it may have an important effect in average organization productivity. Accordingly, the objective of this study is providing a methodology to measure the contribution of worker's stress levels to the firm's overall efficiency. The suggested measurement technique will be provided by a renowned methodology to perform a study on productivity and production efficiency: Data Envelopment Analysis (DEA). The main reason that DEA is preferred is it is a non-parametric approach that avoids technological changes therefore does not require any variable definition for technology (Tyrone, Chia-Chi and Tsui-Fen, 2009). Also, to balance the deterministic environment of DEA, Bootstrap analysis is performed over the production data. The methodologies that are mentioned will be explained in Section 2 in detail. This study distinguishes from other stress-related studies with several notions. To the best of our knowledge, most of the studies about stress in work environment consider human focus, which means main idea is to reduce the work stress for the employees. This study, however, focuses more to the managerial side of the stress as well as employees. Another distinction is that this study contains solely two different manufacturing firms in Turkey. Since stress has a cultural side, perspective of stress is also changing depending where it is occurred.

This study will have several contributions to operations research literature as well as to the worker's health and industrial and organizational psychology literature. For operations research and management, it provides significant insights for the firms to take into account like labor quality process and what impact can occur due to stress.

It will also be supported that DEA is a reliable choice to measure the productivity and production efficiency when latent variables like stress are included. Additionally, one benefit for worker's health and industrial psychology literature is that the worker's stress levels can be identified regarding the outcomes of the study and can be used for the future work on stress in automotive and packing industry. Finally, this study provides insights about the current pandemic and its mental effects over employees since we added a question to WSQ about pandemic.

It is believed that human engineering has drastic effects on firm's sustainability. Sustainable production and balance work environment is something that many organizations lack. Managers are not giving enough attention to worker's stress unless it affects the company's benefits. This is our main motivation to study worker's stress. We would like to take the attention to the worker's stress levels because every individual deserves to be in a peaceful work environment that not only fulfills firm's goals but also the individuals' in this case the employees' expectations from the firm itself. Managers and employers can observe the effects of stress that affect not only the workers but also the firms. And with the aid of this study, companies may go one step further to develop a sustainable, worker friendly work environment.

The thesis will proceed with Section 2 which covers literature review about techniques and methods we used in our study. In Section 3, description of our data and data preparation for DEA and IRT are discussed. Section 4 explains our main methodologies in detail. Analysis and results are given in Section 5 and finally in Section 6, we summarized the findings and discussed the future work.

CHAPTER 2: LITERATURE REVIEW

Literature review contains brief information about stress in the workplace and the conducted experiments. Starting with the DEA which is the main methodology that is used in the study. The studies about production efficiency and productivity consists of two main methodologies: Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA). Although the selection of the methodology might be affected by the data variables and objective of the study in some cases, the selection process can be considered arbitrarily which mostly depends on the preference of the researchers (Wadud and White, 2000). In this study, DEA is used as a main methodology to model the efficiency since it is non-parametric and uses deterministic input and output data. Continuing, Malmquist Productivity Index is used to measure the relative efficiency and productivity through a specific time period. It is one of the most robust techniques to conserve the efficiency changes as well as comparison over two different economic technologies. Since a qualitative variable is considered in the thesis (stress), one of the most crucial parts is to create an accurate and detailed survey to measure the stress levels of the workers. For this purpose, Work Stress Questionnaire (WSQ) is adopted. This survey is designed specifically for measuring the work stress. In Section 2.3, WSQ will be discussed in detail. Review will proceed with Item Response Theory (IRT) to create a better understanding on qualitative survey to quantitative results and will give brief information about Graded Response Theory which is a sub-method in IRT. And also provides brief information about Confirmatory Factor Analysis. Maximum Likelihood Estimation is another methodology that is used to estimate the proper parameters for GRM. The model, its history, advantages and disadvantages will be mentioned as in the previous sections. Chapter 2 continues with Mokken Scale Analysis and its brief review. Another crucial method is bootstrapping which is used to get more realistic results through the given data. Lastly, the program RStudio that is used to analyze and interpret the results is discussed.

This study has several differences compared with the peer studies that also aim to display the contributions of stress to work efficiency. One of the strengths of this study is the usage of Work Stress Questionnaire which is proved to be a renewed and robust measure for stress in work environment. To the best of our knowledge, this has not been used for Turkish firms before.

2.1 Data Envelopment Analysis

In this section, the development of Data Envelopment Analysis (DEA) throughout the years is explained and some crucial studies about the approach are given. Additionally, since this study mainly focuses on production applications, studies about manufacturing are extended.

2.1.1 Data Envelopment Analysis: History and Applications

Farrell (1957) planted the seed of DEA by seeking an efficient method for determining productivity. They did not directly use the basic DEA model (CCR model) but its dual. Therefore, the dual of the CCR model is also known as Farrell Model. Main suggestion of this monumental article was that the current methodologies were too restrictive for the use of multiple inputs that would make the evaluation process closer to the realistic standards (Farrell, 1957). DEA is introduced as a term by Charles, Cooper and Rhodes (1978). It can be considered as a revision on Farrell's study that measures the efficiency of several schools (Charles, Cooper and Rhodes, 1978). Later on, the simplest and most popular DEA model is named as CCR model which is a combination of the initials of three authors.

After the developments and the increased attention about the technique, DEA became a renowned subject for operations research and production studies as well as economics. Energy to environment, tourism to sports, DEA is one of the most commonly used production enhancement methods in both service and production industries. Figure 1 displays the distribution of the studies according to the areas of study. Regarding popularity, since it is quite challenging to mention all areas of study, we focus on five areas: Banking, Health Care, Agriculture, Transportation and Education. Overall, between 1975 and 2010, 10.31% of the studies are about banking, followed by 8.65% health care, 8.23% agriculture, 7.95% transportation and 5.87% education (Lui et al., 2013). The milestones and crucial studies about the top areas will be discussed to create an overall idea about how DEA evolved and affected the literature along the way.

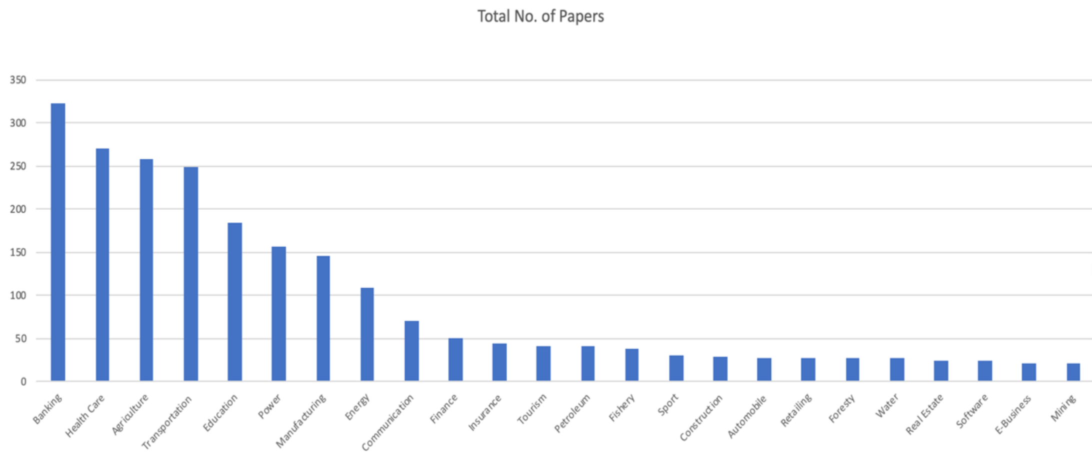


Figure 1. Distribution of the studies regarding the area of study in DEA

Banking: As a field, banking is one of the most crucial areas of study since efficiency is the key measure regarding the development pace. The first use of DEA in the banking industry goes way back to the study of Sherman and Gold (1985). In this article, they applied the DEA to measure the efficiency of bank branches, in those days DEA was a new method to use in any industry so they used the classical CCR model which was getting popular and popular with each and every article published. Sherman and Gold showed that DEA has numerous advantages to provide a managerial insight through efficiency. Additionally, they included the limitation of the method on banking industry as well for the potential future works. Rangan (1988) took Sherman and Gold's work one step further and again studied DEA to measure the technical efficiency in 215 banks. One significant contribution is that Rangan had realistically enough random sample from a sufficient number of banks compared with Sherman and Gold's study. Concluding the article, study provides drastic results which indicates that 70% of the input is enough to produce the same amount of output. Berg, Forsund and Jansen (1992) applied DEA to measure the productivity of three Nordic banks in Norway, Finland and Sweden. The productivity differences are measured by Malmquist indices (bilateral index which is used to compare two different productivity levels of two different economies). They showed that Swedish banks have better position among Nordic banks.

Continuing, DEA evolved and is being used even more in the 21. century with the enhancements of the previous studies and different DEA methods are introduced like network DEA model. Banking efficiency performance measures also became almost default hence a procedure-like method is developed to facilitate and expedite the estimation of productivity and efficiency.

Health-Care: Health-care is also substantial as a field in every literature since it requires rapid response and efficient service to make treatments as effective as possible. An insufficient healthcare system might cause mortalities, which is the main concern of every nation, compared with the properly managed peers. Measuring the efficiency of healthcare systems begins with two significant articles Nunamaker (1983) and Sherman (1984). Nunamaker (1983) studied the efficiency of nursing service efficiency in Wisconsin hospitals and resulted those more inefficient hospitals are identified via DEA compared with the previous studies on the same hospitals. Additionally, Sherman (1984) provided answers for some groups of teaching hospital efficiencies and concluded that two hospitals are inefficient according to the DEA.

Agriculture & Farm: As having a significant role in overall development process, DEA was uniquely new concept for agriculture when Färe, Grabowski and Grosskopf (1985) published their article on Technical Efficiency of Philippine agriculture. In this article, authors developed three different efficiency measures and resulted that even though some flaws can be observed, Philippine agriculture was mostly efficient regarding the provided data. Apart from this, Chavas and Aliber (1993) worked on estimating the different efficiency levels in Wisconsin farms. This article is a distinct article with the property of not only measuring the technical efficiency but also calculating scope, scale and allocative efficiency as well.

Transportation: Schefczyk (1993) analyzed 15 different airlines' efficiencies, regarding their operating cost, available ton km, non-flight assets as inputs and revenue pass. The methodology included the extension of previous methodologies with DEA and resulted that for 14 out of 15 airlines, the focus of core business of passenger transportation has been beneficial to the performance. Taking as an example, proceeding article adjusted the airways efficiency model that is constructed regarding the DEA applications.

Education: As a field, education has been the pioneer for the DEA applications with just starting early with the article that assesses productivity with the aid of mathematical programming (Bessent et al., 1980), which follows the work of Charnes, Cooper and Rhodes and analyzes the productivity in Houston Independent School District and resulted that 46.7% of the schools were inefficient. Additionally, article also highlights the difficulties about data collection.

2.1.2 Data Envelopment Analysis in Manufacturing

2.1.2.1 History

Since DEA was proposed in 1978 by Charnes et al., it has been used by many industries with several different purposes like managerial insights, discovering potential gains, improving production activities etc. To the best of our knowledge, the potential use of DEA for managerial enhancements starts with Norsworthy and Malmquist in 1983. Study mainly analyzes the multi-factor productivity growth in Japan and the USA. Although DEA is not mentioned as a direct methodology, the process they followed through is quite similar to DEA (Norsworthy and Malmquist, 1983). Proceeding with Epstein and Henderson, study displays significant measurement and requirements for diagnosis and control for production efficiency via DEA. Additionally, the article gives substantial insights about the limitations and advantages of the methodology for both in diagnosis and performance measurement (Epstein and Henderson, 1989). The limitations and benefits of DEA on manufacturing application are further discussed in Sections 2.1.2.3 and 2.1.2.4. In the late 90's, DEA was also considered as a significant tool for staffing efficiency in manufacturing which can detect any misinterpretation that affects efficiency (Ward et al., 1997). Entering the millennial, green manufacturing was growing through the literature of production and operations research. In 2005, DEA also gave significant insights for green manufacturing process assessment (Zhang and Wang, 2005). Regarding the wide application field, with the developing technology and the growth in data science, contemporary DEA articles often emphasize the improvement and adjustments of the currently developed model especially for the use of social experiments.

2.1.2.2 Model Development

Although DEA models can vary for numerous applications such as two-stages DEA, BCC models, in manufacturing, adjusted versions of basic CCR models are more often used. Comparing several articles, the main adjustment to the CCR model is the addition of integrity variables to model technology and identifying peer DMUs. Intensity variables defined and adjustments made on model that is considered in the study are discussed in Section 4.2. briefly.

2.1.2.3 Advantages and Strengths

Despite peer methodologies, the main advantage of DEA is that it can handle many input and output models. This brings important variety to the models as well as facilitates the struggle with big data. Additionally, it doesn't require an assumption or specific definition for the objective function since it contains the ratio of the inputs and outputs (Ali and Lerme, 1997). Another critical advantage of DEA is the capability of handling the ordinal and qualitative data. This extends the application field of the DEA with the support of several models such as Graded Response Models. Also, with the dual of the specified DEA model, the behavior of the DMUs can be observed. Lastly, any inefficiency caused by DEA models can be also identified via DEA (Charles, Cooper and Rhodes, 1978).

2.1.2.4 Disadvantages and Limitations

Although DEA has lots of benefits for both researchers and firms to assess the efficiency, it also has some limitations as well. The inputs and outputs must be chosen carefully due to the sensitivity for both inputs and outputs of the model. In addition, high efficiency levels may lead researchers to misinterpret the results when comparing relative efficiencies. DEA is a method that is quite open to errors since it ignores the statistical errors. This leads numerous researchers to use data correction methodologies such as re-sampling or bootstrapped DEA. Proceeding with, DEA can evaluate a specified period of time which means it cannot handle the changes due to time. Conversely, this problem can be eliminated by Malmquist – like DEA indices. Last but not least, DEA ignores the impacts of exterior variables that are also a part of operations.

2.2 Malmquist Productivity Index

In this section, foundation and development of Malmquist Productivity Index (MPI) is mentioned briefly with the extension of manufacturing applications. Also, index development and its interpretation are given as well.

2.2.1 MPI: History and Applications

Malmquist Index was first studied by Stem Malmquist in 1953, which basically represents the amount by which one consumption bundle must be scaled in order to generate the same utility level provided by some base consumption bundle (Grifell-Tatje and Lovell, 1995). The main contribution to the measurement of productivity, however, was developed by Caves et al. (1982) and also supported by Nishimizu and Page (1982) where they expand the MI to study the total factors productivity as well as relative efficiency (Caves, Christensen and Diewert, 1982). Since then, MPI has had a considerable amount of space in operations research literature. Agriculture to banking, airlines to public sectors, MPI has been chosen as a main productivity estimate.

Public Sector: Applications in the public sector have a special place in the literature since DEA is used for the estimation of Malmquist Index for the first time by Fare et. al. (1994) where the authors analyze Swedish hospitals. They used the Malmquist productivity index to measure the technical efficiency and efficiency changes. The results displayed the efficiency changes in 17 hospitals and additional to that technical efficiency levels indicated the improvements as well as the regress (Färe et al., 1994). Additionally, Magnussen (1994) also studied 46 different Norwegian hospitals and again resulted the considerable amount of difference in productivity measures. Again, Fare et. al. (1997) worked on international productivity growth in health care delivery in 19 OECD countries with two different DEA models. Other than the health-care sector, MPI gained popularity through Norwegian road sector studied by Odeck (1993) where they investigated the productivity changes for the time period 1989-1991.

Proceeding with Taskin and Zaim (2000), where they used the non-parametric Malmquist Index model to study the public enterprise sector in Turkey for years between 1974 -1991, they discovered the public sector growth was significantly lower than private sector growth.

Banking: The measurement of banking industry with MPI mostly includes the studies within a specific country which means the international comparisons are not very popular. Again, the applications contain the classical components of MPI which are technical efficiency and efficiency changes. Berg et al. (1992) created an input-based productivity model to observe the efficiency changes in Norwegian banks. As outputs, they used the loans and deposits and resulted that for the first periods the regress occurred and later on progress of the sample average (Berg, Forsund and Jansen, 1992). Similarly, Fukuyama (1995) studied an input-based MPI for Japanese banks and resulted in a uniform decline in productivity. Bauer et. al. (1995) also displayed a similar result that can be computed by input-based model where authors examined the U.S. banks between 1977-1988 and again the outputs were loan based.

Agriculture: Thirtle, Hadley and Townsend (1994) used input-based MPI to estimate and compare the productivity levels and agriculture sector growth in several African countries and resulted that productivity growth is not significant but mostly positive (Thirtle, Hadley and Townsend, 1994). Also, an output-based method was developed by Tauer (1994) in order to estimate the productivity in U.S. dairy farms in three different articles. The article has an important place in the literature since the author proves the adjustability of the MPI with adding chance constraints. Lastly, Ozden, Armagan and Bekcioglu (2010) justified the productivity and efficiency measurement significance around the developing countries such as Turkey with the study they conducted on crop production all around the country and resulted in efficiency changes occurring due to regional differences.

Transportation: McMullan and Okuyama (1996) applied a non-parametric approach in order to estimate MPI to measure the productivity in U.S. motor carriers and found drastic technological decreases occurring in a given time period (McMullen and Okuyama, 1996). Another unique study is measuring the changes of carbon-dioxide emission and its effects on transportation productivity with the aid of Malmquist

Environmental Productivity Index. It is proved that the increase in carbon-dioxide has negative impacts on transportation productivity in the U.S. between the years 2002-2012 (Choi and Roberts, 2015).

2.2.2 MPI: Applications in Manufacturing

Productivity and efficiency changes have a significant effect on manufacturing applications. If it is not done properly, simple statistical analysis can cause thousands of dollars and reputation loss for the organizations as well as the negative impacts on countries' economic growth. Accordingly, Sowlati and Vahid (2006) measure Canada's wood manufacturing sector using the MPI and DEA again with the core component of MPI: technical efficiency and efficiency changes. The results were quite interpretable and showed the drastic improvement on Canadian manufacturing over 8 years between 1994 - 2002 which is caused by mainly frontier shifts (technical efficiency improvements) (Sowlati and Vahid, 2006). Another striking study conducted in Hi-tech industry of China by Qazi and Yulin (2012) who measure Total Factors Productivity with output based MPI and DEA to observe the productivity changes for 15 different hi-tech industry firms and concluded that the office equipment industry is on the lead with 3.7% productivity gain (Qazi and Yulin, 2012). Also, a recent study is conducted on Indian textile manufacturing using output oriented MPI by Gambhir and Sharma (2015). Authors stated significant evidence about the effects of technical efficiency and changes on productivity gain in 160 textile companies irrespective of their scale (changing between small to large segment firms). Interpretation of MPI in manufacturing applications suggests even though input or output-based analysis doesn't affect the outcome much, to facilitate and forecast the behavior of productivity movement, output oriented MPI and DEA are appropriate.

2.2.3 MPI: Index and Interpretation

MPI is a bilateral index which is commonly used for explaining the differences of two economies regarding their production technology. It is derived from a production function which gives technological relations of inputs and outputs. The commonly used model for MPI is to determine the efficiency changes over time. Therefore, it is crucial to understand the main components of MPI which are the catch up or recovery and frontier shifts, in other words the innovation. Catch up as a term refers to which amount a specific DMU can achieve for improving its efficiency. Frontier shift displays the shift in the efficiency frontiers around the decision-making units (DMU) between two different time periods. Accordingly, MPI is the combination of these terms:

$$\text{MPI} = (\text{Catch up}) \times (\text{Frontier shift})$$

Catch up: To examine the MPI in detail, it is necessary to understand the components and how they are calculated in the index. Let $(x_0, y_0)^1$ and $(x_0, y_0)^2$ be the sets in production possibility function for time periods 1 and 2 where x refers to the input vectors and y to the output vectors, which denote the input and outputs as vectoral values. The catch-up effect is calculated as follows:

$$\text{Catch up} = \frac{\text{Efficieny of } (x_0,y_0)^2 \text{ w.r.t. period 2 frontier}}{\text{Efficieny of } (x_0,y_0)^1 \text{ w.r.t. period 1 frontier}}$$

Above ratio can be interpreted as follows: if its value is greater than 1, that shows progress in relative efficiency and otherwise, it means a regress or no change over the given period of time.

Frontier Shift: It is a significant component of MPI in order to evaluate the productivity changes. Let δ_1 be frontier shift for period 1 and δ_2 for period 2. The equation of the frontier shift at $(x_0, y_0)^1$:

$$\text{Frontier Shift Effect}(\delta_1) = \frac{\text{Efficieny of } (x_0,y_0)^1 \text{ w.r.t. period 1 frontier}}{\text{Efficieny of } (x_0,y_0)^1 \text{ w.r.t. period 2 frontier}}$$

Similarly for $(x_0, y_0)^2$:

$$\text{Frontier Shift Effect } (\delta_2) = \frac{\text{Efficieny of } (x_0,y_0)^2 \text{ w.r.t. period 1 frontier}}{\text{Efficieny of } (x_0,y_0)^2 \text{ w.r.t. period 2 frontier}}$$

Using the given equations, Frontier Shift is calculated as the geometric average. If the frontier shift is greater than 1, it shows a progress in the front technology from period 1 to 2; otherwise, it refers to no change or regress (Tone, 2006).

2.3 Work Stress Questionnaire

Work Stress Questionnaire (WSQ) was developed by K. Holmgren et al. in 2009. The aim of the questionnaire is the early detection of the individuals who are at risk of being mentally, physically unhealthy or in a disturbed state of mind due to stress in the workplace (Holmgren, Hensing and Dahlin-Ivanoff, 2009). WSQ contains 21 distinctive questions to measure the stress in the workplace. Each question measures different situations that might cause stress like time to finish the assignments, thinking about work etc. One of the advantages of WSQ is it requires a reasonable time. Since a lot of the survey applications take more than 1 hour, that causes outputs to be biased because individual focuses are distracted over time and answers for the questions become insufficient and sometimes totally irrelevant with the state of the subjects. Another benefit that WSQ provides to the researchers is that, since stress can be affected by work but also can affect the efficiency of work, WSQ comprises both work-related factors and personal characteristics (Holmgren, Fjallstrom-Lundgren and Hensing, 2013). One additional comment on WSQ is that it is a relatively new questionnaire which takes recent situations into account that are affecting individuals and causing them to develop stress.

Since it is a questionnaire that measures stress, WSQ is mostly used in worker health and safety or psychology literature. It is used to measure the work-stress related and its possible impacts on future absenteeism (Holmgren, Fjallstrom-Lundgren and Hensing, 2013). In another significant cohort study, it is used to display the relationship between work-related stress and sick-leave or health issues of employed Swedish women. It is reported that, when the work-related stress levels are high in employed women, the chances are higher for them to be sick-listed or report health issues. Study also reveals the most common types of work-related stress which is work interference with leisure time (Holmgren et al., 2009). Questionnaire is also used to test its validity on male workers by Holmgren-Frantz (2019). Study is performed on male workers aged between 18 to 64 in Sweden and concluded that WSQ is a viable option to measure the stress among male workers and displays the positive correlation between sick leave and work-related stress (Holmgren and Frantz, 2019).

The original revised version of WSQ can be examined in Appendix A. Regarding the WSQ direct translated version to Turkish language is also added to the Appendix B which is discussed in Section 3.

2.4 Item Response Theory

This section provides significant insight about development of Item Response Theory (IRT) and its applications. Also, comparison between Classical Test Theory and Item Response Theory is given which highlights the main reasons that IRT is preferred as one of the main methodologies in our study.

2.4.1 IRT: History and Applications

Item Response Theory (IRT) is a statistical approach specifically studied to evaluate the effects of latent variables that are measured via questionnaires, survey or patient-reported outcomes (PRO's). IRT has a set of psychometric models often used for developing psychological measures. The main motivation for IRT was to create models that directly measure the intangible qualities of mind (Thomas, 2011). Theory first introduced in 1950's by Frederic Lord, Georg Rasch and Paul Lazarsfeld with numerous studies and gained its popularity in 2000 with a detailed study done by Embretson & Reese and since then became mainstream in many disciplines especially in psychology and health-care.

One of the most common uses of IRT is evaluating the changes caused by the personal differences, like which age group, gender, ethics have impacts over developing psychological variables such as developing a panic disorder; therefore, IRT is also a psychometric tool. One important study contributes that direction of the item keying has no significant effect over measuring personality and psychopathology to the psychology literature (Reise and Waller, 2003). Also, mindfulness is an increasing subject in many societies. IRT studies also includes the analysis of mindfulness with

Mindful Attention Awareness Scale (MAAS) (Van Dam, Earleywine and Borders, 2010).

2.4.2 IRT: Advantages and Limitations: Comparison with CTT

Although its release is relatively new compared with the peer theories like Classical Test Theory (CTT), the pros and cons are quite clear with the studies performed throughout the years. For instance, in CTT, it is challenging to compare different traits of individuals at the same psychological variable continuum. However, with IRT, it is possible to get latent estimates from a sample with different types of individuals (Reise, Ainsworth and Haviland, 2005). IRT also outperforms study performed on leadership skills; it shows that IRT is a viable option for estimating significant variables in human resource management (Reeve, 2002). Additionally, CTT has strict assumptions on reliability measures such as Cronbach's α (which is a significant measure for consistency that estimates how closely related the items in a given set of samples) are distributed normally and equally for all grade levels, which is not valid compared with the real world. On the other hand, IRT assigns different precision levels for each item and basically explains that precision is the least for a specific item where items do not discriminate well (Reeve, 2002). Proceeding, for CTT to be reliable, there is a minimum level of scales, in other words, to have a viable scale to measure a specific latent variable it must be longer. IRT eliminates the necessity with shorter scales; it can also be valid and equally viable. In addition, CTT reliability measures are also sample dependent which means some values that are generated may not be able to lead different research since it's specific for that unique sample. Compared with CTT, IRT measures are sample-invariant (Reeve, 2002). Despite its advantages, one main limitation that research reported on IRT is the model complexity and variety. Secondly, most IRT models need larger sample sizes than CTT differentiating between 500-1000 (Reeve, 2002). However, it is proved that in the Graded Response Model which is an IRT model even with small sample sizes it gives interpretable outcomes. Table 1 summarizes the comparison and displays some other advantages of IRT over CTT.

Table 1. Comparison of IRT over CTT

Classical Test Theory	Item Response Theory
Measures of precision fixed for all scores	Precision measures vary across scores
Longer scales increase reliability	Shorter, targeted scales can be equally reliable
Test properties are sample dependent	Test properties are sample free
Mixed item formats lead to unbalanced impact on total test scores	Easily handles mixed item formats.
Comparing respondents requires parallel scales	Different scales can be placed on a common metric
Summed scores are on ordinal scale	Scores on interval scale
	Graphical tools for item and scale analysis

2.5 Graded Response Model

This section gives important insights about what Graded Response Model is and its development through time, proceeding with benefits and disadvantages with proofs to highlight the proper application of the method.

2.5.1 GRM: History and Applications

Graded Response Model (GRM) include several mathematical models in Item Response Theory, which are designed to express the qualitative polygamous categorical latent factors as quantitative values. These categories may have several option types such as, letter grading and states (disagree, agree, somewhat agree etc.) for given questions (Samejima, 1997). It is first introduced by Samejima (1969), who developed the model with inspiration of latent structure analysis and created a monograph. The model has three main uses: Most frequent use is to determine the probability of which grade or score tests subjects might receive. It is used quite often to create an insight for university course schedules, for firms to evaluate the customer responses and forecast the customer segment accordingly. Another use is to estimate the subject's latent trait. GRM can be a facilitating model to understand the effects of latent variables. For instance, Gummelt, Anestis and Carbonell (2012) examined the Levenson Self Report Psychopathy Scale (LSRP) with GRM to create a link between assessment of the psychopathic individuals and their personality traits. The items of the LSRP's are grouped in order to create a latent trait and with GRM, and the frequency of these traits is determined to observe the effects of the gender differences (Gummelt, Anestis and Carbonell, 2012). Processing main uses, GRM can also display the power of the test items. Alias, how well the test items can evaluate the ability or latent trait such as research questionnaires can be validated by GRM. While translating the questionnaire from its original language a validation is usually needed in order to get realistic results. Stănculescu (2021) validates Romanian version the Fear of COVID-19 Scale (FCV-19S) by the aid of GRM (Stănculescu, 2021).

2.5.2 GRM: Advantages and Limitation

It is stated that GRM is beneficial for numerous application areas especially health-based questionnaire research due to the fact that sample requirement is more flexible compared with the peer methodologies. Even with small sample sizes, GRM can be quite informative (Depaoli, Tiemensma and Felt, 2018). Of course, it also depends on the complexity of the GRM if it is adjusted for a specific use. In other words, for some complex GRM's, large sample size might refer to more than thousands of responses. Additionally, GRM creates a fitter model in determining the statistical parameters compared to the Classical Test Theory in human resource research (Siengthai and Sukirno, 2010). Despite the advantages, the main limitation of IRT models, especially GRM, is that the mathematical models that are developed are really complicated compared to the peer models. However, this is also slightly eliminated by the programs like STATA and R since they automatically give results of the GRM with the algorithms that are already in the system or appear as an extension of the programs.

2.6 Confirmatory Factor Analysis

This section explains Confirmatory Factor Analysis (CFA). It gives brief information about the methodology and highlights the advantages and disadvantages. Additionally, comparison with Exploratory Factor Analysis (EFA) is included to give insights about the renowned techniques about factor analysis.

2.6.1 CFA: History and Applications

CFA is first introduced with the technique of basically any parameter can be fixed at any level and the remaining parameters may be calculated by the maximum likelihood estimation (Jöreskog,1969). According to the author, procedures (both exploratory and confirmatory analysis) eliminate several different problems that might occur during previous factor analysis method. Methodology itself is derived from common factor model which is the pioneer study on factor analysis that concludes factor scores are dependent with both selected variables to be measured and the individuals that are chosen for the study (Thurstone, 1947). Since its development, both confirmatory and exploratory factor analysis were in the research scene and aided numerous studies especially psychometrics and statistical analysis of latent variables.

Most of the studies that uses CFA is in major of psychology. It is quite effective and facilitating for researchers in almost any area of psychology. For instance, Schwartz and Boehnke (2004) explained the human values in 27 different countries with the support of CFA. Recently, Javelle et al. (2021) used CFA to create the German Three-Factor Impulsivity index (TFI) to explain the emotion-related impulsivity.

2.6.2 CFA: Theory and Interpretation

CFA is a structural equation model which is used to estimate the relationship between indicators and latent variables. For factor analysis, crucial point is to determine the proper parameter values. In almost all CFA models there are three parameters, which are factor loadings, unique variances and factor variances. Other parameters can be defined by the researcher depending on the objective of the study. Before modeling specifications and analysis, these parameters must be classified as either fixed, free or constrained (Brown and Moore, 2012).

Proceeding with model specifications, CFA model can be treated as a linear regression model with slight differences. For the sake of simplicity, we give a diagram to summarize the approach in Figure 2.

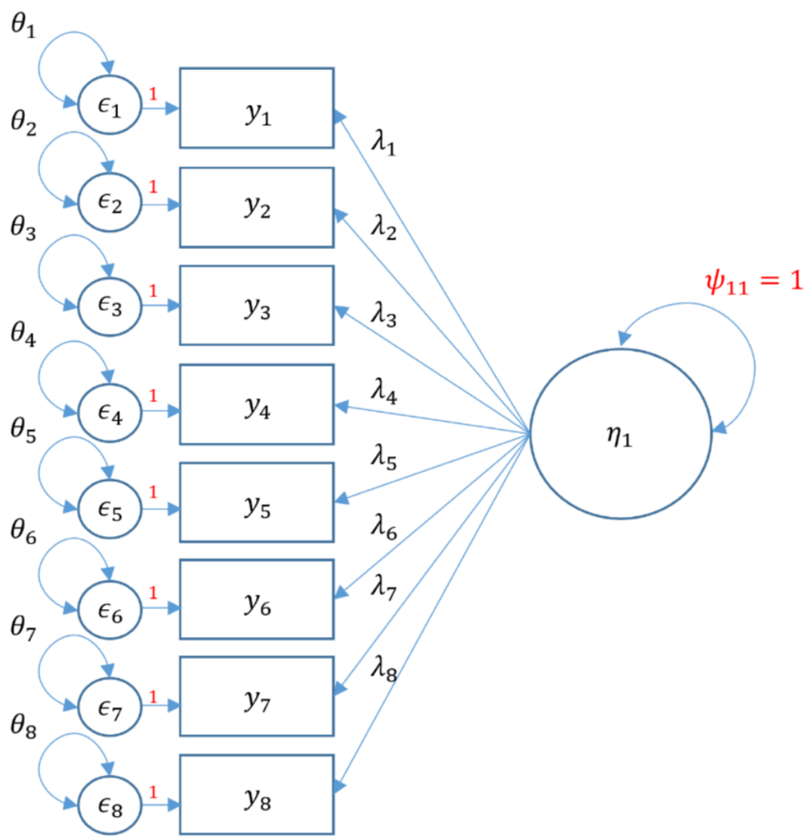


Figure 2. One factor CFA with more than three items

Figure 2 summarizes the CFA theory with one factor (η_1) and 8 items (y_i) where θ_i represents the variance-covariance matrix of residuals of factor model, ε_i corresponds to the residuals in other words latent factors that certain questionnaire cannot explain, λ_i shows the loadings or weight of the items that can be interpreted as the correlation of each item with the latent trait or factor and ψ_{11} is the variance-covariance of the factor, which is set to 1 in this case since the factor scores for the items are the aim of the study. This means that we are not interested in trying to find the factors; rather we have a factor that is predetermined and the objective is to observe how we can explain the factor with the given questionnaire (UCLA: Statistical Consulting Group, 2021).

2.6.3 CFA: Comparison with EFA

The main difference between CFA and EFA is that in CFA models it is crucial to describe a priori or model which explains the model structure; however, EFA is mainly used to determine the factor structure which isn't predetermined. Another important difference between two techniques is in EFA we don't have a certain knowledge of what factors can be affected by our variables; we might have an idea of it, but it is not directly known. In CFA, there is a known factor or factors that are desired to be measured by defined variables. There is a hypothesis about what is actually expected from a certain factor to affect the estimated measures (Suhr, 2006).

2.7 Maximum Likelihood Estimation

In this section, MLE development throughout the years is explained and some of the important studies about the method is given. Additionally, statistical model derivation and its interpretation is mentioned. Finally, advantages and disadvantages are mentioned briefly.

2.7.1 MLE: History and Applications

Maximum Likelihood Estimation (MLE) was first introduced by Fisher in 1922 and developed further by the author in 1925 with regarding the consistency and efficiency, and since then, it is a widely used parameter estimation methodology in statistical inference. Fisher also studied the significant properties of MLE and later on Cramér and Rao studied the theory in detail. The relation of MLE and consistency and efficiency are mentioned by Fisher. However, the development and detailing are made by Cramér (1946). The discussions about consistency and efficiency continued almost thirty years by several authors. Rao (1960) supported that there is no evidence for MLE to be widely applicable for different statistical models since there isn't enough proof regarding the consistency which is also concurred by Le Cam (1960). Neyman and Scott (1948) proved that although MLE is generally consistent for some cases, i.e. when the number of parameters is increased, it might lack consistency. The discussions on efficiency were quite similar with consistency. It is stated that, for a limited class of estimators, MLE is a good technique regarding efficiency. Although, the criticism may direct MLE might be lacking in some circumstances, it is still an optimal parameter estimation methodology that can be applied to many different probabilistic models like regression models, graded response models etc. (Norden, 1972).

2.7.2 MLE: Model and Interpretation

Estimation process begins with determining the likelihood function of a given data regarding the model parameters. Each data set follows a different probability distribution with the changes of parameter values and accordingly a probability density function. Probability density function (pdf) is defined in the form of an integral of the density of the variable density over a given range and can be displayed as given in Equation (1).

$$P\ (a\ \leq X \leq b) = \int_{a}^{b} f_X\ (x)\ dx \tag{1}$$

where *X* is a random variable and $f_x(x)$ is an integrable probability function of *x*.

Regarding the pdf, it is possible to create a likelihood function and process with the estimation over it. Pdf displays that some observations are more likely than the other ones in the given sample. When we have data (which in this case we already know the data and its properties), likelihood function aims to find a pdf that is most likely to produce the data which is studied. In this case, likelihood function can be defined as in Equation (2).

$$L(w|y) = f\,(y|x) \tag{2}$$

where *L* represents the likelihood function for data vector *y* and parameter vector *w*.

According to the likelihood function, the main objective for the parameter estimation is finding a value for the parameter vector that maximizes the likelihood function. MLE estimates sometimes do not exist and in order to find a proper estimate, model parameters might require different constraints to get an optimal estimate (Myung, 2003).

2.7.3 MLE: Advantages and Disadvantages

When the model is properly assumed and its parameters are strictly defined, MLE is one of the most efficient estimators compared to the peer estimator models like Least Squares Regression and Generalized Method of Moments. It is also beneficial when a large sample is studied since it can provide unbiased estimates. Additionally, sometimes some assumptions that are made over a model must be violated in order to have the desired measures for the study. MLE estimates are consistent and reliable even in such cases.

Along with the benefits, MLE can be sensitive for the starting values like any other optimization methods. As it is quite useful in larger samples, small samples might generate biased estimates regarding the model. Lastly, in some cases it might be quite challenging to create the likelihood function when the model derivation is complex.

2.8 Mokken Scale Analysis

In this section, we explain the basics of Mokken Scale Analysis (MSA), beginning with the history and common application areas and proceeding with the theory and approach briefly.

2.8.1 MSA: History and Applications

Mokken Scale Analysis is first introduced in 1971 by political scientist Rob Mokken, where he focused on nonparametric dichotomous IRT models and item scores. Since then, it's one of the most common procedures used by research for data reduction and questionnaire scaling, especially for social science studies. Developments over MSA started to rise in 1991 with its application over polytomous item scores (Molenaar, 1991 and 1997). Also, in the early 2000's, studies also covered parametric IRT model and item scores as well (Meijer and Baneke, 2004). To facilitate the use of MSA, the most popular MSA commercial software was developed in 2002 under the name of Mokken Scale Procedure (MSP) (Molenaar and Sijtsma, 2002). However, since its release there is no update or no future updates are anticipated, therefore, it is not reliable to perform the procedure via MSP (Van der Ark, 2012). One of the most useful software packages is developed between 2007-2010 for RStudio as mokken. This package is mainly designed to conduct MSA over RStudio and now it is the most commonly used software package in the area (Van der Ark, 2007).

2.8.2 MSA: Theory and Approach

Mokken Scale Analysis is a psychometric technique for scaling an ordinal data. Its theory is developed from nonparametric IRT models. Since it is a derivative form a nonparametric IRT model, MSA aids researchers to investigate if all the particular rules defined for IRT model are held. Hence, MSA is one of the most popular methods for data reduction as well (Van der ark, 2012).

Theory of MSA is very much alike with IRT models with slight differences. It would be redundant and excessive to mention all the mathematical representations and models since this isn't our main methodology rather performed mainly for data reduction. However, one formula that is substantial to understand the concept is for calculation of item scalability coefficients. Let H_j be the item scalability coefficient for item j, item scalability coefficient for item j is calculated as follows:

$$H_j = \frac{COV(X_j,R_j)}{COV\left(X_j,R_j\right)^{max}}$$

where X_j is the item score for item j and R_j represents the rest score. Additionally, the scalability of the test is calculated similarly as well:

$$H_j = \frac{\sum_{j=1}^{J} COV(X_j,R_j)}{\sum_{j=1}^{J} COV\left(X_j,R_j\right)^{max}}$$

2.9 Bootstrapping

This section provides insights about development of Bootstrapping and its applications. Also, foundation of the theory and how to apply the approach for a certain data is mentioned briefly. At last, the limitations and benefits of the approach is explained.

2.9.1 Bootstrapping: History

It is first introduced by Efron (1979) where he mentioned it as for large amounts of computation in the place of traditional mathematical models to construct sample liability measures like confidence intervals, variance etc. In his article he also states that methodology especially aids for complex analysis like multivariate analysis . Later on, Singh (1981) made significant comments on Efron's bootstrap, comparing with the peer methodology jackknife and bootstrap and also analyzed the convergence of the bootstrapping method. Same year, a Bayesian bootstrap was developed by Rubin. Additionally, Efron (1987) examined the technique for better confidence intervals and its bias correction. Since its release, bootstrapping has become one of the most common statistical techniques in resampling methods due to the ability of application for any statistical measure to estimate the sample properties.

2.9.2 Bootstrapping: Theory and Approach

Bootstrapping is a methodology that creates realistic samples for a specified deterministic sample via mimicking the sampling process. The approach is basically generating a sample (resampling) from a given sample for predicting the population and its properties. Even though the population is unknown, bootstrapping technique assumes it is known which is the given sample. Subsequently, it is measurable in that sense. From a bootstrap distribution many statistical parameters estimate can be found like confidence intervals. It is not only for the parameter estimation but also for eliminating the error that can be caused by samples behaviors. Generating a bootstrap distribution for the given sample might aid the studies to converge to realistic results even using deterministic data rather than stochastic one.

There are several bootstrap methodologies depending on the objective and sample properties. For instance, block bootstrap is used to improve the accuracy of the bootstrap for time-series data, the wild bootstrap is studied in the context of regression models with heteroskedastic variables (Davidson and Flachaire, 2008). Therefore, it is crucial to determine the goal of the bootstrap to get an accurate result.

2.9.3 Bootstrapping: Advantages and Limitations

The main advantage that makes bootstrapping popular in statistics is that it is a quite straightforward and easy method. Additionally, it is significantly useful when the distribution of the data is unknown or not accurately estimated, bootstraps regenerate a sample data from a given sample so that the data is more accurate and ready for valid analysis. Along with the advantages bootstrapping technique has one important limitation. For rare extreme values bootstrapping method may fail by ignoring them, which causes it to generate a data that is worse than the original one (Ebert, 2018).

2.10 R Studio

In this section, History of RStudio and its applications are given briefly. Studies are given to prove how versatile the program might become. Also since the main methodologies for our study is DEA and IRT, some functions and packages specifically assigned for the methodologies are mentioned as well.

2.10.1 R Studio: History and Applications

RStudio is a code-based statistics tool with many features that enables to perform almost all data analysis steps one by one. In December 2010, RStudio started to developed and released in early 2011. It is written in both C++ and Java but mostly in JavaScript. Therefore, RStudio is a versatile tool that provides users various programming languages to work with including one of the most useful language Python.

RStudio has more than 10,000 packages that each has unique features. Subsequently, it is used in every area for many different objectives. For instance, in very recent article that is publish in June 2021 explores the zakat administration in times of Covid-19 via text mining in Indonesia using RStudio. The study has theological side as well as its contribution on health-care (Hudaefi, Caraka and Wahid, 2021). This is only a one of various example that RStudio can be used in various fields.

2.10.2 R Studio: DEA and IRT

DEA is a very renowned method amongst the other frontier analysis methods. Therefore, RStudio developers provided frontier analysis packages like benchmarking and FEAR. Ødegaard and Roos (2014) used FEAR package in R to measure the technical efficiency and the contribution of psychosocial work environment and worker's health to efficiency levels. Wilson (2008) explains the FEAR package and its advantages in detail where they give crucial tips and insights for functions and its arguments in FEAR. Another way to perform DEA on R is using benchmarking library. However, benchmarking is not as versatile and flexible as FEAR, since it does not provide significant adjusted analysis like various compositions of Malmquist index (Wilson, 2008).

For IRT analysis to get latent trait behavior, again R is a powerful tool with providing different libraries such as mirt (maximum likelihood estimation for IRT models), ltm (latent trait models), grm (graded response models) regarding the data type and objective of the study.

CHAPTER 3: DATA ANALYSIS

Study is conducted for two different firms (firm names can't be provided due to privacy issues) and for each, data collection process is divided into two parts. One for applying the WSQ (Work Stress Questionnaire) to employees to observe the latent trait pattern which in this case is stress. Original questionnaire includes 35 questions; however, with the authorization of the developers, regarding the current situation, 2 questions that measure the stress due to pandemic are included. Appendix A displays the original version and Appendix B includes Turkish version with the questions about the current pandemic included.

The other collection part is to determine the technical efficiency and efficiency changes via Malmquist-like productivity index and DEA (Data Envelopment Analysis). Each of the processes will be mentioned separately for each firm in detail.

3.1 Firm 1

In this section, Data properties of Firm 1 is mentioned in detailed. Proceeding with, survey arrangement, bootstrap and ppm value calculation is explained.

3.1.1 Data Properties

Firm 1 is an international company which mainly produces ignition coils for automobiles. It has several plants within 5 continents. Our primary focus however, is the plant located in Izmir. As in all frontier analysis methods, input and output values must be determined accurately. For firm1, first stage of data collection is performed via hard copied questionnaires. Respondents' names are not shared but their age, gender and education status has been collected in order to understand the variety. The data consists of 50 responses from different employees aged between 25 and 40, who are only white collar.

For the second stage, required input and output values for efficiency analysis are provided by the firm as it can be seen in Tables 2 and 3, respectively. These values are classified into two different categories as input and output values. For input, production days, number of workers and total work hours in the given month has been collected. It can be observed that production days and total work hours are very similar in input level regarding what they indicate. This might cause some error on DEA in certain situations. However, in this study these 2 input values don't conflict, therefore it doesn't affect the outcome of the DEA model. This conclusion has been made after checking the model with and without production days as an input value.

Table 2. Input values of Firm 1

		INPUTS		
Year	**Month**	**Production Days**	**Total Work Hours**	**# of Workers**
2020	**September**	30	675	273
	October	31	697.5	270
	November	30	675	269
	December	30	652.5	272
2021	**January**	31	697.5	273
	February	28	630	271
	March	31	697.5	273

Table 3. Output values of Firm 1

		OUTPUTS		
Year	**Month**	**Production Amount**	**Quality Index (QI)**	**QI (adjusted)**
2020	**September**	2068447	15190.93	8103.89
	October	1827684	30414.13	7788.72
	November	2095066	8973.79	10411.3
	December	2189549	7788.72	30414.13
2021	**January**	1673253	10411.30	8973.79
	February	2209804	8103.89	15190.93
	March	1981459	9411.29	9411.29

For output, production amounts and quality indices (PPM values) has been provided. Data covers a period between September 2020 and March 2021. Months are chosen with the assumption of current pandemic might also have an effect which has seasonality over the period of time where it doesn't get prevalent in summer but recorded cases are increasing after August 2020. We assume that pandemic might have a significant impact over worker stress so including the data where pandemic effects are drastically different might cause misinterpretation.

3.1.2. Data Analysis

Data arrangement is the most crucial part of the study since working with psychometric analysis requires a considerable amount of work. Additionally, since deterministic data is used, bootstrap analysis is performed via Excel over production amounts values to get accurate and realistic results.

3.1.2.1 Survey Arrangement

After finishing the survey process, the results are documented to an Excel file. Since "yes or no" and "yes, maybe or no" type questions do not directly measure the stress levels, those questions are removed to reduce the number of questions and get more accurate results via Item Response Model. This means that only questions that has 4 options (not stressful, little stressful, stressful, very stressful) and questions that measure individuals' indirect stress factors (Question 1, 2, 3, 4, 19, 20, 21) are included in the IRT model. After eliminating, options are numerically coded between 0 and 3. Question with higher score indicates that an individual perceives the given situation more stressful than employees with lower score. One important note is for Questions 19, 20 and 21; the numerical coding is done reversely because for these questions, first option indicates more stress and the last means less. Last adjustment for Firm1's WSQ data is made to get feasible results from IRT model. Normally, for polytomous categories GRM (Graded Response Models) is applied to understand the response patterns. However, it requires a significant participation, preferably over 200. Since our sample is not enough for GRM, dichotomous IRT with 2 parameters is applied. Therefore, options are reduced to 2 as 0 and 1. 0 represents not stressful and little stressful, while 1 represents stressful and very stressful.

3.1.2.2 Production Data Arrangement and Bootstrap

Arrangement over Firm 1's production data consists mainly of the bootstrap analysis and PPM (Parts Per Million) assignment. Bootstrap is a powerful analysis for deterministic data in statistics. As it is mentioned before, it is applied via Excel using *RANDBETWEEN ()* function. The main motivation to apply bootstrap is that the production amount is significantly high for every month. Higher amount of deterministic data has a higher chance of misleading the study as well as production itself. Additionally, bootstrapping provides more accurate and probable data when it is applied. In this study, we generated 3000 bootstrapped production amounts randomly between the highest and the lowest production levels regarding 5 different production lines for each month. In Table 4, the original production levels are displayed. For instance, the highest production in September is in Line 102 with 695,500 units and the lowest is 134,300 units. So, each bootstrap data will be within these numbers. When it is done, we have 3000 bootstrapped sets, each containing 5 different production amounts representing the 5 lines that produce ignition coils. For each set, the production amounts are summed which means we have 3000 different bootstrapped production amounts. Each value represents a possible production amount. Therefore, we calculated the arithmetic average for these values to get a one month's bootstrapped production amount.

Table 4. Actual production amounts provided by Firm 1

YEAR	MONTHS	Line 108	Line 102	Line 106	Line 103	Line 101	TOTAL
2020	9	134300	695500	660200	406200	459800	2356000
	10	44600	726100	633900	373500	800	1778900
	11	122200	715350	677300	644600	350050	2509500
	12	161900	673500	716500	368800	549700	2470400
2021	1	97100	573500	552500	249400	246200	1718700
	2	191600	616500	694900	350930	379800	2233730
	3	122775	666050	647500	217750	406050	2060125

Final adjustment is made over the quality index values. These values are not directly provided by the company due to privacy issues. However, the company provided graphical representations of PPM values for each month so the values are roughly estimated from the graphs given in Figure 3. Normally, there are 10 different graphs representing each production line but the company allowed us to show one of them to explain how the quality index values are calculated.

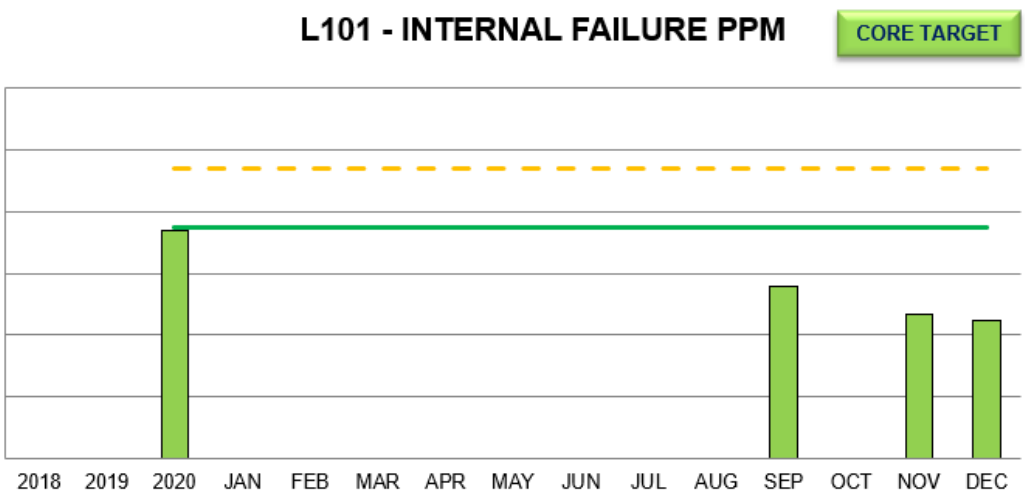


Figure 3. Firm 1 PPM graph

Y axis indicates total defective items per month and each horizontal line represents 1000 units cumulatively. Green line depicts the tolerable limit for units and the intermittent yellow line represents the non-tolerable zone. Preferably, values should not exceed the green line; however, the zone between yellow and green lines can also be saved by extra work until it exceeds the yellow line. PPM is calculated by Equation (3) with the aid of the graphs where we get approximate estimates for the total number of defective items.

$$\text{PPM} = \frac{\text{Total number of defective items}}{\text{Total number of production}} \text{ x } 10^6 \qquad (2)$$

For instance, Line 101 has roughly 2,200 defective units in January and the total production for that line is 246,200. Regarding Equation (3), PPM value is calculated for Line 101 in January. To calculate the total PPM value for a specific month, each of the graphs are interpreted and an approximate total number of defective items are estimated and PPM values are calculated. Finally, to get PPM values, an arithmetic average is found. In Table 5, the assigned PPM values regarding the graphs is shown under the quality index. One final adjustment over quality index levels is that since PPM values represent the defective parts per million the higher values will represent a worse score; however, our output-based DEA model considers a higher value as a good indication for an efficiency level. Therefore, these values are ranked from highest to lowest and the values are reversed as the highest PPM value becomes the lowest PPM value and vice versa. The adjusted values are also displayed in Table 5 under the QI (adjusted) title.

Table 5. Quality Index Values

Quality Index (QI)	**QI(adjusted)**
15190.93	8103.89
30414.13	7788.72
8973.79	10411.3
7788.72	30414.13
10411.30	8973.79
8103.89	15190.93
9411.29	9411.29

3.2 Firm 2

In this section, Data properties of Firm 2 is mentioned in detailed. Proceeding with, survey arrangement, bootstrap methodology done via Excel is explained.

3.2.1. Data Properties

Firm 2 is a company that primarily focuses on plastic injection and flexible packaging. It has 2 production plants in Izmir, Turkey. For this company, there is no direct product in a single line, since they have rapidly changing production. Therefore, our focus is more on how much kilogram of products are produced in a certain period of time rather than what type of product and number of products is created. For firm 2, first stage of data collection again starts with questionnaires which are collected via hard copied questionnaires as well as in digital platforms like Google forms, survey monkey and emails. Firm 2 confirmed to us that participation rate is almost 100%. 212 employees have participated out of roughly 230, who are aged between 23 – 55; and respondents include both white and blue collar.

As production data, input and output values for DEA are provided by the firm as it can be seen in Tables 6 and 7, respectively. These values are classified into two different categories as input and output values. For input, number of workers and total work hours in the given month has been collected.

Table 6. Input values for Firm 2

		INPUTS	
Year	**Months**	**Total Work Hours**	**# of Workers**
2020	**September**	675	273
	October	697.5	270
	November	675	269
	December	652.5	272
2021	**January**	697.5	273
	February	630	271
	March	697.5	273

For output, production amounts and quality indices has been provided. Unlike Firm 1, quality indices are not PPM values, rather they are direct percentages of how much kg wasted from a certain batch of input. Lastly, for Firm 2, data covers the same dates as Firm 1 which is September 2020 to March 2021 due to possible impact of pandemic.

Table 7. Output values for Firm 2

		OUTPUTS		
Year	**Months**	**Production Amount**	**Quality Index (QI)**	**QI(adjusted)**
2020	**September**	490226.2	10.80	10.68
	October	441186.3	11.41	9.79
	November	434122.0	11.27	10.54
	December	466130.5	10.83	10.8
2021	**January**	344215.8	10.68	10.83
	February	445693.7	10.54	11.27
	March	441527.8	9.79	11.41

3.2.2. Data Analysis

Unlike Firm 1, Firm 2 has a slightly larger sample to analyze for survey. This allows us to perform a more sophisticated and accurate method for analysis which is GRM, a polytomous IRT model. Hence, arrangement process has considerable differences with Firm 1. On the other hand, quality index arrangement is less of a stunt.

3.2.2.1 Survey Arrangement

Since Firm 2 has a relatively larger sample size, gathering sufficient data took approximately 3 months. Again, results are documented to an Excel file. As for Firm 1, "yes or no" and "yes, maybe or no" type questions are removed to reduce the number of questions and get more accurate results via Graded Response Model. This means that only questions that have 4 options (not stressful, little stressful, stressful, very stressful) and questions that measure individuals' indirect stress factors (Question 1,2,3,4, 19, 20, 21) are included. Options are numerically coded between 0 and 3. Question with higher score indicates that an individual perceives the given situation more stressful than employees with lower score. One important note for Questions 19, 20 and 21 is that the numerical coding is done reversely, because for these questions, the first option indicates more stress and the last means less. Since firm 2 has enough sample size to apply GRM which is a polytomous IRT model, binary coding is redundant. Although option coding suitable to be remained as 0 to 3, to get more accurate results options are coded between 1 and 3 (1 = not stressful, 2 = stressful, 3 = very stressful).

3.2.2.2 Production Data Arrangement and Bootstrap

Firm 2 has considerably high production rate regarding the number of products. These products are produced from the same raw materials in 2 lines coded as RG110101 and RG110106. To put it more clearly, a raw material can be used to produce several different products, therefore the company provided the production amount and quality index values in weight. This means that the production amounts shown in Tables 7 and 8 are expressed in kilograms. Unlike Firm 1, Firm 2 directly provided the quality index values as a percentage of scrap rates. Again, high scrap rate means less efficiency hence these values are ranked between months as highest to lowest. For instance, in October the scrap rate is the highest with 11.41% and lowest is March with 9.79%. It means to have a proper DEA model the quality index percentage for March will be 11.41 and for October 9.79. Last but not least, the same bootstrap procedure is also performed to production amounts for Firm 2 with the motivation of proper measurement. Actual production levels are displayed in Table 8.

Table 8. Actual production amount for Firm 2

YEAR	MONTHS	RG110101	RG110106	TOTAL
2020	9	194848	295402	490251
	10	180232	260740	440972
	11	218436	215672	434107
	12	223766	242403	466168
2021	1	193899	149914	343813
	2	228249	217426	445675
	3	256811	184515	441327

CHAPTER 4: METHODOLOGY

Methodology contains brief information about how the study is conducted and the models are applied. Unique distinction of this methodology is: For different firms we performed different IRT to highlight the dichotomous and polytomous IRT models. Also, to the best of our knowledge, there does not exist a study in Turkey that considers stress effects over production efficiency and productivity. Analysis begun with getting factor scores via RStudio using *ltm* and *mokken* library. As it is mentioned previously, Firm 1 does not have enough sample size to conduct an accurate GRM therefore dichotomous IRT is used (Jiang, Wang and Weiss, 2016). On the other hand, Firm 2 provided a considerable amount of participation for the survey so GRM is an accurate model to be applied. After getting different factor scores averages for each firm, DEA model is developed. First, we found the technical efficiency (TE) which is the efficiency levels where stress is not included and then we rerun the model with compiling the average factor scores of stress levels to the technical efficiency model to get stress affected efficiency measures. In order to find stress, effect over efficiency, SEP is calculated. Finally, multiplying these two values (SEP and TE) will reveal the actual stress effect in efficiency levels (Ødegaard and Roos, 2014). These steps will be explained further in detail. Additionally, Figure 4 illustrates and summarizes the analysis.

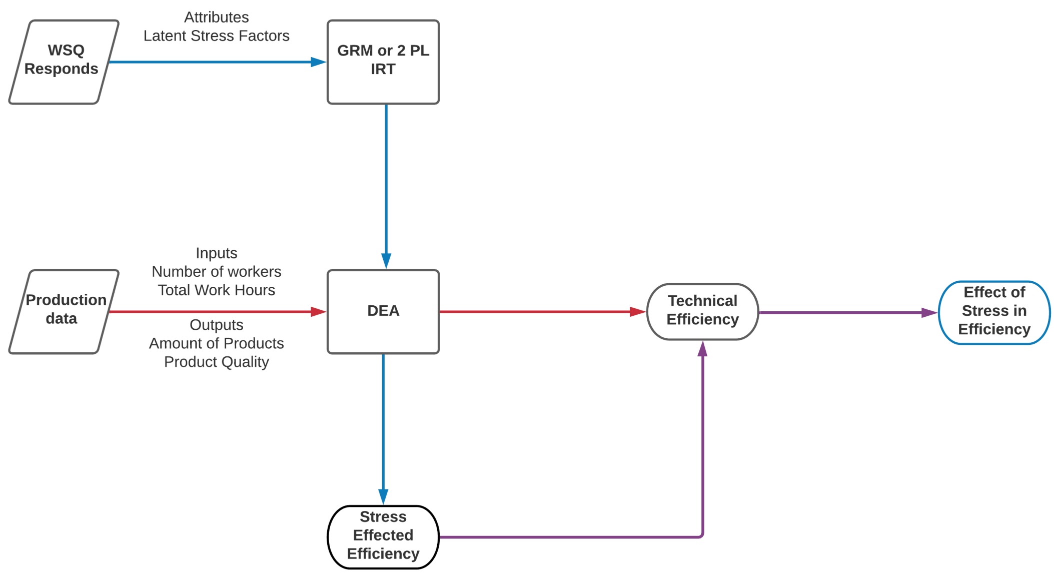


Figure 4. Flow chart of the methodology

4.1 IRT and GRM

IRT is basically used to calculate the probability of which score a certain test subject, latent trait or ability can get. In this study, two different IRT models are used: 2 parameter binary (dichotomous) IRT model and GRM. Mathematically representing, let $P_i(\theta_j)$ be j test participant probability with ability response θ_j the item i right or desired. Equation (4) represents 2 PL IRT model.

$$P_i(\theta_j) = \frac{e^{D_{a_i}(\theta_j - b_i)}}{1+e^{D_{a_i}(\theta_j - b_i)}} \tag{4}$$

In Equation (4), b_i is a constant for the test item i. It is referred as the location parameter, category boundary or item difficulty.

This parameter specifically tells us where the graph is located regarding the standard normal distribution. The location parameter or item difficulty is for determining how high on a latent trait an individual is before they adopt a score.

D and a_i represent the slope parameter together where D is the scale factor of the parameter and a_i can be considered as a discrimination parameter. Discrimination parameter displays the frequency of how well item differentiates between the subjects' responds' score on a specific latent trait or ability.

GRM is also a sub model in IRT so it is also explained by Equation (4) with a slight difference of calculating the probability of specific grade or category. Let i correspond to the items in the questionnaire and j be the category or score index. In Equation (5), GRM shows the probability of the compound events $X_i > j$ (Van der Linden, 2005).

$$P_{ij}(\theta)=\begin{cases} 1 & \text{for } j=1 \\ \Pr\{X_i \geq j \mid \theta\} & \text{for } j=2,\ldots.., m_i \\ 0 & \text{for } j>m_i \end{cases} \tag{5}$$

With these models, the main objective is to get a notion about the pattern of the questionnaire response to calculate the factor score via CFA.

4.2 DEA: Technical Efficiency Estimates

Data Envelopment Analysis (DEA) is a widely used method in operations research and economics to measure the relative efficiency of decision-making units (DMU). It's a data-oriented, non-parametric approach that measures the performances of the DMUs which refers to any entity that has potential to convert inputs into outputs. DEA was first introduced by Charnes, Cooper and Rhodes and the basic DEA model is named after them as CCR (ratio) model. Model below elucidates the CCR model which is the basic model of DEA literature.

CCR model

$$\max h_0(u,v)=\sum_r u_r y_{r0} \Big/ \sum_i v_i x_{i0} \tag{6}$$

subject to :

$$\sum_r u_r y_{rj} \Big/ \sum_i v_i x_{ij} \leq 1 \qquad \text{for } j=1,\dots,n \tag{7}$$

$$u_r v_i \geq 0 \qquad \forall \text{ i and r} \tag{8}$$

In the above model, u_r and v_i are the variables, which can also be mentioned as weights for the input x_{i0} and output y_{r0} for DMU_0. Without the constraints, the model will be unbounded as expected. Therefore, adding a set of normalizing constraints for each DMU is necessary. First, Constraint (7) ensures that the ratio of output and input for every DMU is less than or equal to 1. It does not allow the objective function to exceed 1 since it is a ratio that must have a maximum reference value which is 1. Constraint (8), shows that the weight or variables must have a positive value, to prevent an infeasible solution for the given model. Also note that the model assumes parameters *x* and *y* are also positive since they exist in the model.

The model can be constructed in two different ways regarding the main objective of the study. The model can be either output or input based. Output based DEA models focus on increasing the efficiency by increasing the amount of output levels, on the other hand, input-based DEA models focus on decreasing the input levels to increase the efficiency manufacturing increasing the output levels for efficiency changes is mostly desirable since it means more profit and can lead a company to develop. Rajasekar and Deo (2014) give significant evidence with their study they exercised on major ports in India, which has data that comprises the years between 1993-2013. They elucidate the efficiency levels are approximately the same with both input and output-based DEA models and the evaluation process is not affected drastically (Rajasekar and Deo, 2014).

One crucial adjustment for the CCR model is to add a variable called intensity variable. Intensity variables are used to model returns to scale and identify peer DMU. It can be observed in almost every manufacturing efficiency measurement model with different forms and in constraints depending on the methodology used. In DEA, intensity variables explain the effects of the exact amount of DMUs that participate in objective function. Intensity variables were modeled in the constant returns to scale, throughout the years it is also developed that variable returns to scale can be also modeled by constraining the intensity variables sum into one (Afriat, 1972). The model can also be extended as adding an upper bound on the intensity variables in input constraints. Also, the fractional non-linear model is converted to a linear model with the adjustments.

Regarding the CCR model, DEA model for this study was constructed as an output-based DEA model since our aim is to measure efficiency change via optimal output increase. It is also assumed that an increase in input has proportional increase in output levels as well (constant returns to scale (CRS)) which is the common assumption for many firms. Although, DEA is technically an approach rather than a specific model, representing mathematical model for the output-based CRS-DEA is explained in the model below.

Output based CRS-DEA model

$$max\ \theta \tag{9}$$

subject to:

$$\theta y_{j_m} \leq \sum_j z_j y_{j_m} \quad \forall\, m \tag{10}$$

$$\sum_j z_j x_{j_n} \leq x_{j_n} \quad \forall\, n \tag{11}$$

$$\sum_j z_j = 1 \tag{12}$$

The main objective is to observe how much y_{jm} can be increased. Accordingly, For, m number of output and n number of input, θ represents output increase. Constraint (10) represents the output restriction, Constraint (11) for input restrictions where z_j defines the weight or intensity variables for DMU j and Constraint (12) guarantees total weight is equal to 1. Model provides the most efficient way to use input for maximum level of output. However, in order to get a value between 0 and 1 for output-based efficiency estimates, reciprocal values must be calculated. Hence, *TE* is estimated as given in Equation (13).

$$\mathrm{TE} = \frac{1}{\theta} \tag{13}$$

Regarding DEA, analysis consists of two models that have slight differences in input. In the first model that will be examined, we run the model with 3 input values, which are total hours of work, number of workers and production days and 2 output values as production amount and quality indices. The second model consists of 4 input values as the input attribute stress factor is included and the other entities remain constant. Both models have been constructed via RStudio using *benchmarking* package and compared after interpreting separately.

After estimating the TE levels with 2 different models, the effects of stress are calculated via Equation (14). It represents the indirect stress effect over productivity and efficiency estimates and to facilitate the understanding it is named as Stress Effected Process (SEP) or Stress Effected Efficiency (SEE).

$$\mathrm{SEP} = \frac{\mathrm{TE}}{\mathrm{TE\ (Stress\ Included)}} \tag{14}$$

Productivity and Efficiency Changes (PC) is the product of both TE and SEP so final analysis is to calculate to see any improvement might occur due to production and labor stress levels.

CHAPTER 5: ANALYSIS AND RESULTS

This chapter provides detailed analysis performed over the data provided by both firms. Starting with Firm 1, the IRT results are reported, and proceeding with DEA over the production data and its bootstrap analysis and also with the stress factors and production data combined are discussed. Same procedure is also applied to Firm 2 with slight differences in model to get factor scores.

5.1 Firm 1

In this section, Firm 1's analysis results will be discussed starting with IRT results and factor scores. Then, Firm 1's DEA results with validations and final comments are made to the *TE, SEP* and *PC* values.

5.1.1 IRT Results

As it is mentioned in Section 3.1.2.1, some items have been removed and the model inputs are counted as 22 variables which in this case left over questionnaire items for direct indication of factor scores. Before getting the factor scores, it is crucial to observe if data has any discriminating item. This refers for the items that may not provide significant evidence for the stress level due to low level of response variety. To get the items that might cause a miscalculation over factor scores, 2 parameters dichotomous IRT model has been run to get the parameter values via Maximum Likelihood Estimation. The purpose of using 2 parameter is; it is needed to observe the discrimination parameter (also known as slope parameter) however, in 1 parameter IRT models (Rasch models), only difficulty parameter is calculated which in this case doesn't give sufficient information about the variety of the data. Table 9 displays both the difficulty parameter and discrimination parameter values for each item calculated via *mirt* and *ltm* library in RStudio.

Table 9. Parameter values for dichotomous IRT

	Difficulty Parameter	Discrimination Parameter
Item 1	0.644	21.98
Item 2	-0.606	-0.582
Item 3	4.260	0.807
Item 4	1.380	1.938
Item 5b	-0.076	1.261
Item 6b	0.801	1.185
Item 7b	1.300	1.121
Item 8b	6.662	0.305
Item 9b	0.088	1.295
Item 10b	1.578	0.729
Item 11b	1.102	1.375
Item 12b	0.326	2.615
Item13b	0.937	1.560
Item 14b	1.027	1.850
Item 15b	0.682	38.51
Item 16b	0.394	2.339
Item 17b	1.369	0.990
Item 18b	1.026	2.258
Item 19	0.698	39.73
Item 20	0.659	39.85
Item 21	0.539	2.068
Item 22b	0.170	1.283

There are 22 items to determine the latent variable stress. However, as it is calculated by MLE, some parameter values don't show significant impact on latent variables which will cause them not to fit the item distribution later on. The desired values for discrimination (slope) parameter for a 2PL IRT model is around 1 or 2. In that case items 2 and 8b elucidate a very little discrimination. It indicates that these items are not able to measure the factor scores since they don't have enough discrimination (variation) regarding the responses. Items 1, 15b, 19, and 20 are also removed due to high level of discriminability. High levels of discriminability mostly happen due to very few combinations of answers, where these combinations are significantly different from each other. Although difficulty parameters are also displayed, since the questionnaire used in this study is a subjective survey and doesn't focus on measuring a test difficulty or how well an item can be done correctly, any values that are assigned by the model can be acceptable. Additionally, Combined Item

Characteristics Curve (ICC) has been displayed in Figure 5. ICC shows the relationship between ability and response probability for each item. It is desired to be an S shaped function which states the higher probability means higher ability exponentially.

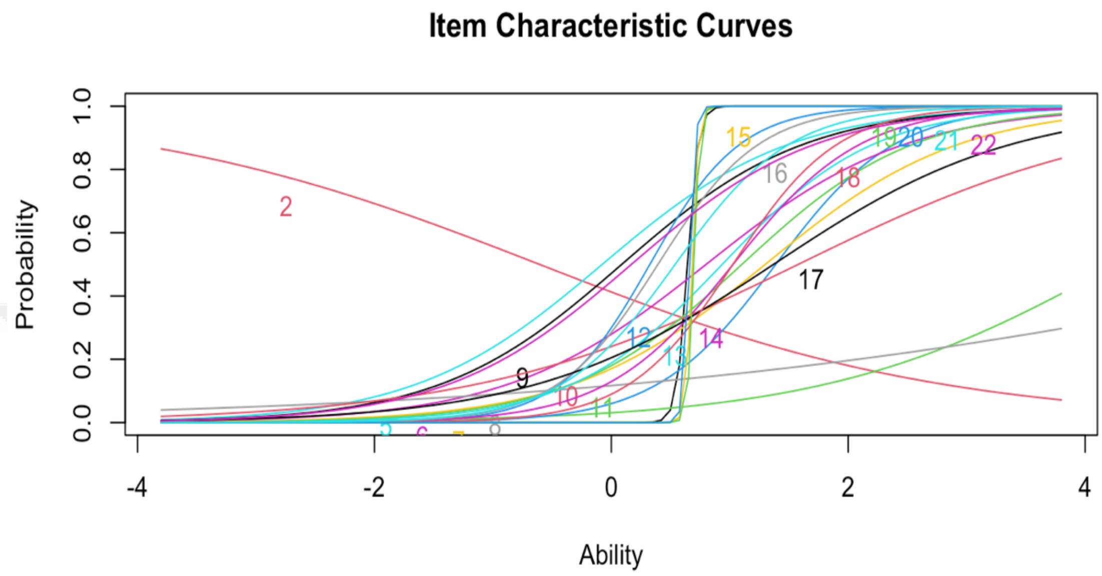


Figure 5. Item characteristics curves for all items

Figure 5 displays that the removed items do not fit the specified rules. For Instance, Item 2 has decreasing trend that indicates low level of ability on estimation in factor scores.

After removing 6 items, the model has been run again with 16 items that are predicted as a suitable indicator. Table 10 proves that these items are a good fit to get factor scores over.

Table 10. Parameter values of remaining items

	Difficulty Parameter	Discrimination Parameter
Item 3	2.504	1.707
Item 4	1.402	2.004
Item 5b	-0.079	1.141
Item 6b	0.706	1.497
Item 7b	0.964	2.208
Item 9b	0.086	1.780
Item 10b	1.161	1.108
Item 11b	1.018	1.615
Item 12b	0.337	2.158
Item 13b	0.901	1.716
Item 14b	0.936	2.390
Item 16b	0.382	2.558
Item 17b	1.255	1.135
Item 18b	1.153	1.784
Item 21	0.692	1.247
Item 22b	0.175	1.266

As it can be observed, item that has the highest discrimination value is item 16b with 2.558 where the lowest value is Item 10b with a value of 1.108. Again, to illustrate the difference between these two models, ICC is created with remaining items in Figure 6.

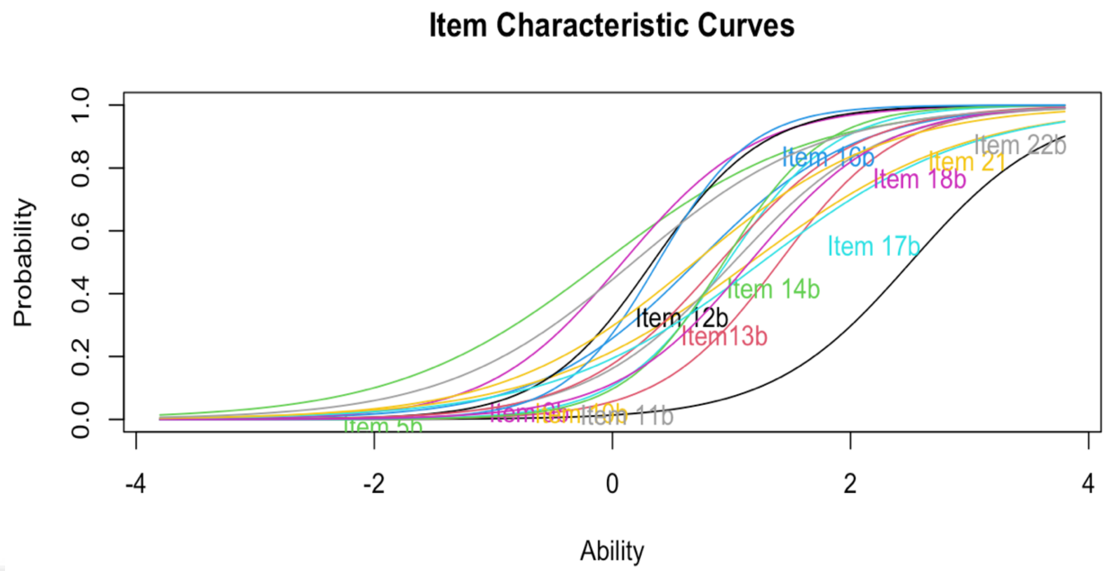


Figure 6. Item characteristics curve of remaining items

The curve can be interpreted as after the elimination of the specified items, all the items have the desired probability and its corresponding ability to measure the latent factor (stress). Also, to make sure every item is fitted to the distribution, the same model is developed over the data, item fit test is run using *item.fit* function in RStudio. Table 11 indicates that there is no significant evidence that 2PL model is not suitable for the data. Hence, it is concluded that 2PL model is a good fit for the data provided by Firm 1. Note that, item fit values are usually biased by number of participants.

Table 11. Item fit results

	X^2	$Pr(>X^2)$
Item 3	6.55	0.37
Item 4	5.63	0.56
Item 5b	6.20	0.80
Item 6b	12.07	0.15
Item 7b	7.94	0.35
Item 9b	7.96	0.48
Item 10b	21.28	0.19
Item 11b	11.11	0.21
Item 12b	4.86	0.71
Item 13b	0.34	0.34
Item 14b	8.10	0.31
Item 16b	8.07	0.29
Item 17b	9.71	0.40
Item 18b	10.80	0.23
Item 21	3.55	0.93
Item 22b	9.16	0.38

After testing the item fit, next step for the analysis over IRT model is to get the Test Information Curve to observe the performance of the whole test. This can be done individually for each item, where it is called Item Information Curves (IIC). However, since our main focus is not examining each item individually, it is created for the whole test.

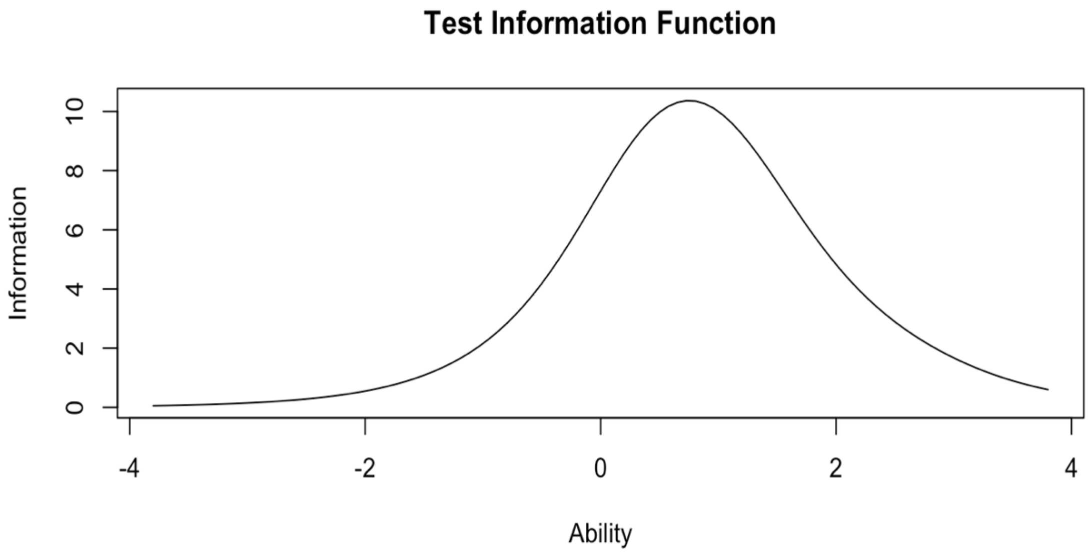


Figure 7. Test Information function of WSQ regarding Firm 1 results

Figure 7 is basically the sum of the all-item information curves for each item. It indicates that the test can show more information on little more than average ability levels. The information provided starts decreasing with the higher level of ability and the least information is provided by the lowest ability level.

Last but not least, to test internal consistency, Cronbach's alpha values have been calculated for each model (before and after reducing the number of questions). It is the most common test reliability measure for psychometric analysis where it tells us how related the items are in a certain questionnaire. In another words, it basically explains if all items can measure the same type of latent variable. For most analyses, its desired value is between 0.70 and 0.90. Accordingly, Cronbach's alpha values for 16 items is 0.87 which is in the feasible zone. Summarized statistics for each item's Cronbach's alpha before the items are dropped can be observed in Table 12, which displays that the reliability levels are roughly the same with the dropped items (the values that elucidate the reliability if an item is dropped is approximately between 0.86 and 0.88).

Table 12. Cronbach's Alpha values before items are dropped

	RawAlpha	Stdized. Alpha	Avrg. R	AlphaSe
Item 1	0.86	0.86	0.23	0.027
Item 2	0.88	0.88	0.26	0.023
Item 3	0.88	0.88	0.26	0.024
Item 4	0.87	0.87	0.24	0.025
Item 5b	0.88	0.87	0.25	0.025
Item 6b	0.87	0.87	0.24	0.025
Item 7b	0.87	0.87	0.24	0.025
Item 8b	0.88	0.88	0.26	0.024
Item 9b	0.87	0.87	0.24	0.026
Item 10b	0.88	0.87	0.25	0.024
Item 11b	0.87	0.87	0.24	0.025
Item 12b	0.87	0.87	0.23	0.027
Item 13b	0.87	0.87	0.24	0.026
Item 14b	0.87	0.87	0.24	0.026
Item 15b	0.87	0.87	0.23	0.027
Item 16b	0.87	0.86	0.23	0.027
Item 17b	0.88	0.87	0.25	0.025
Item18b	0.87	0.87	0.24	0.026
Item 19	0.87	0.87	0.24	0.026
Item 20	0.87	0.87	0.24	0.027
Item 21	0.87	0.87	0.24	0.026
Item 22	0.87	0.87	0.24	0.025
Average:	***0.872***	***0.870***	***0.242***	***0.026***

Raw R value is given under the Item Statistics in Table 13, which shows the correlation between a specific item and total score. This value, however, is biased since the correlation is performed including the item itself so it is not a surprise that according to raw R values almost all items are perfectly correlated.

Table 13. Item Statistics

	N	Raw R	Stdized. Alpha	R correlated	R Dropped	Mean	Std. Dev.
Item 1	50	0.76	0.77	0.77	0.72	0.30	0.46
Item 2	50	0.24	0.21	0.16	0.14	0.58	0.50
Item 3	50	0.22	0.28	0.25	0.45	0.04	0.20
Item 4	50	0.50	0.51	0.50	0.44	0.14	0.35
Item 5b	50	0.47	0.44	0.42	0.39	0.52	0.50
Item 6b	50	0.49	0.51	0.50	0.42	0.32	0.47
Item 7b	50	0.51	0.54	0.52	0.45	0.22	0.42
Item 8b	50	0.21	0.23	0.20	0.15	0.12	0.33
Item 9b	50	0.59	0.58	0.57	0.52	0.48	0.50
Item10b	50	0.40	0.41	0.40	0.33	0.26	0.44
Item11b	50	0.52	0.52	0.51	0.45	0.24	0.43
Item12b	50	0.68	0.67	0.65	0.62	0.40	0.49
Item13b	50	0.57	0.57	0.55	0.50	0.26	0.44
Item14b	50	0.61	0.62	0.61	0.55	0.22	0.42
Item15b	50	0.70	0.68	0.68	0.65	0.24	0.43
Item16b	50	0.68	0.68	0.67	0.63	0.38	0.49
Item17b	50	0.44	0.46	0.43	0.37	0.24	0.43
Item18b	50	0.57	0.58	0.57	0.52	0.20	0.40
Item 19	50	0.64	0.62	0.63	0.59	0.20	0.40
Item 20	50	0.68	0.65	0.66	0.63	0.30	0.46
Item 21	50	0.57	0.55	0.54	0.50	0.34	0.48
Item 22	50	0.49	0.49	0.46	0.41	0.46	0.50

To overcome possible errors, R correlated and R dropped values are also calculated by alpha function in RStudio. R dropped is basically drops the item so that the correlation analysis is done unbiasedly. If R Dropped value is less than 0.3, it can be concluded that the items are correlated with scale overall. For instance, there are two values 0.14 and 0.15 which belongs to Item 2 and Item 8b. It means that these items are not properly correlated with the overall scale.

5.1.2 Factor Scores

Factor score analysis is a crucial part of this study, since it explains the properties of the latent trait that is desired to be measured. Also, it is required to observe how latent traits have an impact over a certain type of statistics. The idea of calculating factor scores is to assess a value for each combination that might occur in the questionnaire with using Confirmatory Factor Analysis (CFA). In RStudio, the library *ltm* already uses CFA in IRT model to calculate the factor analysis so there is no need to create an additional model to estimate the factor scores but the *factor.scores* function in RStudio. Analysis indicates that 8 out of 50 individuals perceive their workplace as non-stressful and a considerable number of employees think minimum amount of stress exists. Only a few of the participants decided that they have stress caused by work. Mathematically exemplification, the interpretation of factor scores regarding the approach is, individuals who decided that their work environment is non-stressful have an ability estimate of -1.251 (8 of 50 individual responses). Appendix C represents the possible combination of their occurrence frequency (Obs.), assigned factor scores for each combination (z1) and standardized factor scores (se.z1). For each combination, there is a unique z1 value, i.e. factor scores. These scores aid the study to determine a value for the stress levels so DEA model can get as an input. One final calculation done on factor scores is to get the mean score so that DEA model can read the average stress factors for the productivity and production efficiency determination. The arithmetic average of the factor scores is roughly 0.278.

Finally, to justify that the factor scores are compatible with 2PL model density curve is created as it is given in Figure 8. It is normally distributed as expected. So, it can be concluded that factor scores are estimated properly by the 2 PL IRT model.

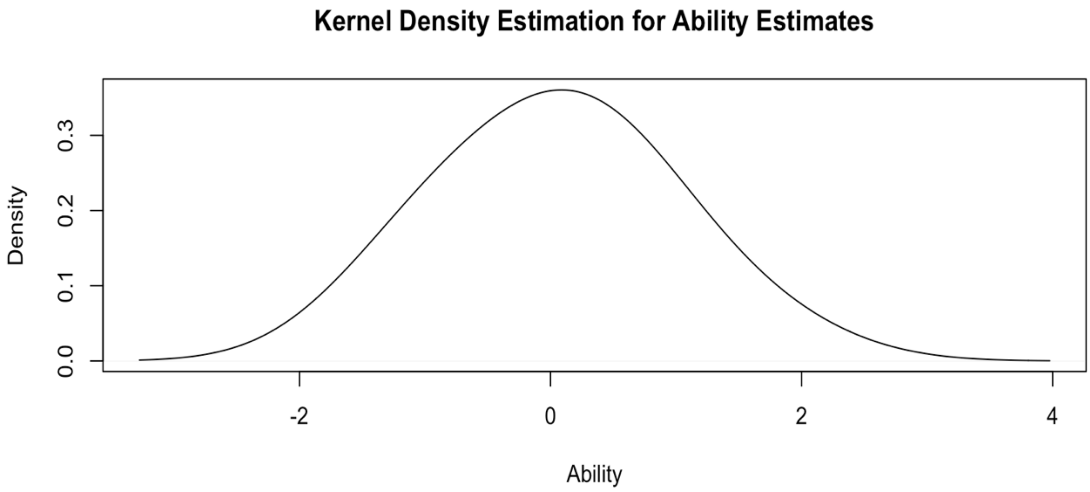


Figure 8. Density curve shows the compatibility of factor scores with 2PL IRT model

It is normally distributed as expected. So, it can be concluded that factor scores are estimated properly by the 2 PL IRT model.

5.1.3 Technical Efficiency (TE)

The term technical efficiency refers to the ability of a system to create maximum output with the minimum required input. Before calculating the efficiency levels for each month, Shapiro-Wilk test is applied to see if the data is normally distributed. The *p*-value has been estimated as 0.2885 which is greater than chosen alpha level (0.05) indicates that tested data is normally distributed. To illustrate the general idea of frontier analysis, Figure 9 is created where y1 represents the efficiency level and y2 represents the production levels also here the circles that are close to the frontier line can be interpreted to be more efficient compared to the other months. Note that Figure 9 is created regarding the output-based DEA. In RStudio, output-based efficiencies are between 1 and infinity so the values may outlie from the feasible area. As it is mentioned before, the output-based efficiency levels are between 1 and infinity. Hence, the reciprocal values are estimated.

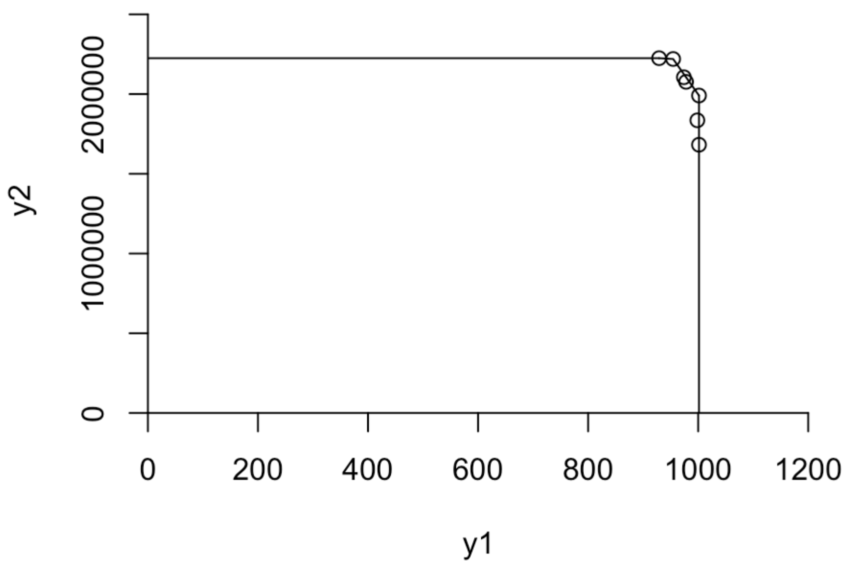


Figure 9. Efficiency graph created in RStudio

Proceeding, Tables 14 and 15 represent the output-based DEA results for Firm 1. For December and February, the efficiency is estimated as 1 which shows that the optimal production is reached in terms of output-input balance and January is the least efficient month with values of 0.751.

Table 14. Technical Efficiency Levels 2020

	2020			
	September	**October**	**November**	**December**
Eff.	1.076	1.205	1.047	1.000
1/Eff.	0.929	0.830	0.955	1.000

Table 15. Technical Efficiency Levels 2021

	2021		
	January	**February**	**March**
Eff.	1.330	1.000	1.123
1/Eff.	0.752	1.000	0.890

5.1.4 Stress Effected Process (SEP)

Stress effected process (SEP) represents the efficiency levels of labor process specifically affected by stress. In the previous chapter, factor scores are calculated in order to observe the stress pattern and its possible effect over production efficiency. Again, prior to calculating SEP, Shapiro-Wilk test is applied and resulted that the data is normally distributed with a *p*-value of 0.2506. Onwards, the efficiency levels are recalculated in Tables 16 and 17. These values are not the direct indicates for SEP; rather they represent merely a measure of TE with stress. In order to find the SEP, Equation (14) is applied and combined results are depicted in Table 18. Results elucidate that Firm 1 stress levels are close to perfect balance and Firm 1 can be considered to have a non-stressful work environment.

Table 16. Technical Efficiency with Stress (2020)

	2020			
	September	**October**	**November**	**December**
Eff.	1.068	1.205	1.047	1.000
1/Eff.	0.936	0.830	0.955	1.000

Table 17. Technical Efficiency with Stress (2021)

	2021		
	January	**February**	**March**
Eff.	1.321	1.000	1.115
1/Eff.	0.757	1.000	0.897

Table 18. Stress Effected Process (SEP) values for Firm 1

	September	**October**	**November**	**December**
SEP	0.99	1.00	1.00	1.00

	January	**February**	**March**
SEP	0.99	1.00	0.99

5.1.5 Productivity Changes (PC)

PC represents the productivity changes of firms which has possibility to be affected by both SEP and TE (Johansen, 1968). This concept is used in this study in order to depict the effects of both production efficiency and stress effects on firm overall productivity separately. PC value is simply calculated by multiplying the SEP and TE indices. Table 19 displays the calculated PC values.

Table 19. Combined values of SEP-TE and PC values

	SEP	TE	PC
September	0.990	0.929	0.920
October	1.000	0.830	0.830
November	1.000	0.955	0.955
December	1.000	1.000	1.000
January	0.990	0.752	0.744
February	1.000	1.000	1.000
March	0.990	0.890	0.881

These values provide significant evidence that SEP has almost no effect on overall firm productivity compared with TE. Of course, this result is expected since Firm 1 has a non-stressful work environment as it was concluded previously. Figure 10 is also created to show the dependency of PC.

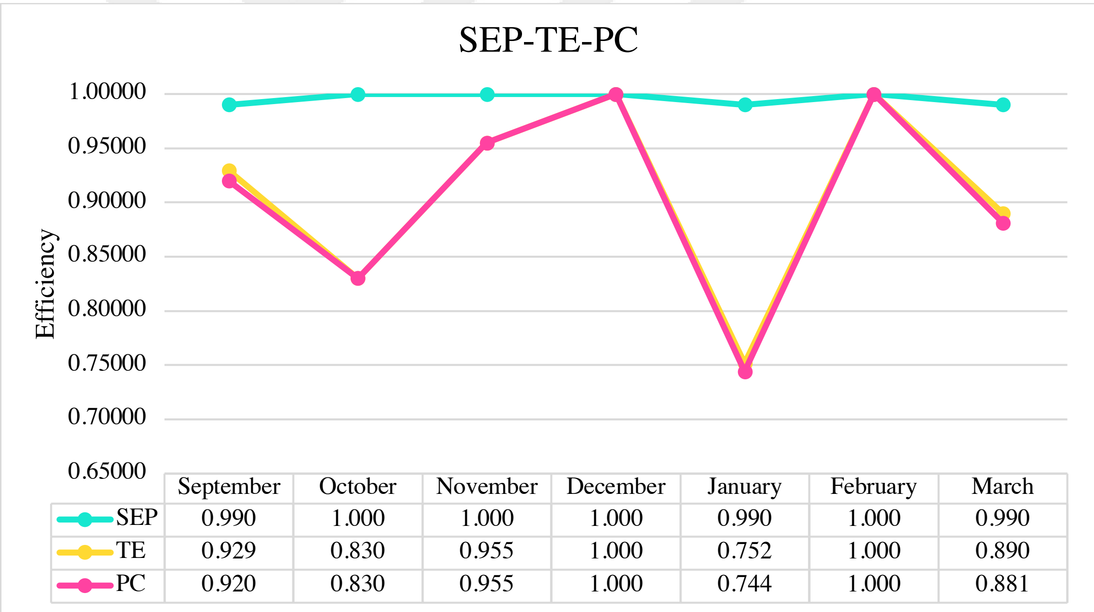


Figure 10. Showing correlation between SEP, TE and PC

5.2 Firm 2

This section starts with MSA and GRM results for Firm 2 and continues with factor score estimation. We finalize the section with efficiency estimates gathered by DEA and TE, SEP and PC interpreted.

5.2.1 MSA and GRM Results

With remaining 22 items from WSQ, the GRM model has been run to observe if the items are proper estimates for latent trait (stress). Substantial part here is to determine whether keeping discrimination parameter constant or not. Therefore, we run 2 different models. In the first model, discrimination parameter is kept constant, and in the second model, it is varying across items. Beginning the interpretation with the first model, Table 20 shows the parameter values calculated by GRM.

Here we have 3 different parameter values: extremity parameter 1, extremity parameter 2 and discrimination parameter (constant). Extremity parameters basically tell us the latent ability score needed by a respondent to have a 50% chance of selecting a specific option. For example, extremity parameter 1 for item 1 means that level of ability score which in this case -1.153, that there is a 50% chance of selecting the option 1 over option 2 or option 3. Note that, if there are *k* categories to be able to select in an item, number of extremity parameters is equal to *k*-1. In this case, since we have 3 options in our questionnaire, model generated 2 different extremity parameters.

Table 20. Discrimination parameter kept constant

Coefficients:	Extrmt1	Extrmt2	Dscrmn
Item 1	-1.153	0.93	1.02
Item 2	-1.599	0.833	1.02
Item 3	-1.366	0.616	1.02
Item 4	-1.471	0.861	1.02
Item 5b	-1.664	0.643	1.02
Item 6b	-1.561	0.616	1.02
Item 7b	-1.569	0.865	1.02
Item 8b	-1.775	0.953	1.02
Item 9b	-1.694	0.705	1.02
Item 10b	-1.543	0.849	1.02
Item 11b	-1.338	1.11	1.02
Item 12b	-1.446	0.913	1.02
Item 13b	-1.738	0.947	1.02
Item 14b	-1.654	0.514	1.02
Item15b	-1.766	0.826	1.02
Item 16b	-1.655	0.703	1.02
Item 17b	-1.621	0.828	1.02
Item 18b	-1.728	0.639	1.02
Item 19	-1.887	0.673	1.02
Item 20	-2.055	0.56	1.02
Item 21	-1.904	0.95	1.02
Item 22b	-1.511	0.373	1.02

Also, Table 21 indicates the varying discrimination parameter values calculated by GRM.

Table 21. Discrimination parameter is varying

Coefficients:	Extrmt1	Extrmt2	Dscrmn
Item 1	-1.401	1.108	0.801
Item 2	-1.552	0.809	1.067
Item 3	-1.2	0.55	1.225
Item 4	-1.36	0.799	1.136
Item 5b	-1.931	0.736	0.84
Item 6b	-1.362	0.544	1.251
Item 7b	-1.476	0.816	1.113
Item 8b	-1.732	0.928	1.056
Item 9b	-1.613	0.672	1.096
Item 10b	-1.484	0.816	1.077
Item 11b	-1.365	1.121	0.998
Item 12b	-1.327	0.841	1.155
Item 13b	-1.827	0.988	0.955
Item 14b	-1.584	0.494	1.084
Item15b	-1.566	0.738	1.217
Item 16b	-1.557	0.665	1.117
Item 17b	-1.533	0.785	1.104
Item 18b	-1.533	0.571	1.216
Item 19	-2.2	0.771	0.835
Item 20	-2.144	0.579	0.965
Item 21	-2.199	1.083	0.845
Item 22b	-2.732	0.607	0.52

Our first objective is to determine which model is the best regarding their measurement of latent trait. Accordingly, ANOVA is applied. Results show that model 2 does not indicate a significant improvement for the measurement of latent results (ANOVA p-value = 0.291, which is not statistically significant). Therefore, we will continue our analysis using the constrained model.

Second step is to eliminate the items that don't accurately measure the latent trait(stress). Unlike, dichotomous IRT, explanation of item availability for measurement of latent trait is not very transparent to get from parameter values or ICC. Hence, in order to create a proper measurement, Mokken Scale Analysis is performed. This analysis mainly clusters or in other terms scaling the items regarding their measurement of the specific scale. This will also explain the relationship between stress and items. MSA begins with the decision of item clustering. There are several ways for clustering the item in a scale. However, the most renowned and easy way is using Automated Item Selection Procedure (AISP) which mainly uses the maximum likelihood-like methodology to create a scale. Fortunately, RStudio's *mokken* library has *aisp* function to directly generate the created scales via AISP.

Table 22. AISP Results

ITEMS	SCALE
Item 1	0
Item 2	1
Item 3	1
Item 4	1
Item 5b	0
Item 6b	2
Item 7b	2
Item 8b	0
Item 9b	1
Item 10b	0
Item 11b	2
Item 12b	1
Item 13b	0
Item 14b	1
Item 15b	2
Item 16b	2
Item 17b	2
Item 18b	1
Item 19	3
Item 20	3
Item 21	0
Item 22b	0

As it can be seen from Table 22, there are 3 different scales according to the analysis. The items that are in the same cluster or scale has significantly high potential to measure the latent trait together. According to AISP, 7 items belong to scale 1, 6 items are in scale 2, 2 items are in scale 3 and 7 items are not scaling at all or in other words in any of the scales that doesn't give substantial measures. (The items which scale values are assigned as 0.)

Since we have one latent trait to measure, we don't have to decide which scale is the best measure for stress. Rather, decision is simply based on the scale that has the highest number of items, that is scale 1 with 7 items.

To ensure that created scale is rational and properly generated, scalability coefficients of the scale are calculated. Table 23 shows scalability values for each item in the scale.

Table 23. Scalability values

	Scalability Values (H)
Item 2	0.317
Item 3	0.325
Item 4	0.307
Item 9b	0.33
Item 12b	0.329
Item 14b	0.315
Item 18b	0.327

These values basically show whether these items are correctly placed in the current scale. There are no discriminating values in the scale, and additionally, overall scalability value for our new scale is roughly 0.321, which indicates that the scale that is tested is a robust scale to test desired latent trait. Also, we tested the monotonicity of the scale and there are no critical values, meaning that all items show monotonicity. Finally, to ascertain the reliability of the scale, Cronbach's alvpha value is calculated as 0.812. This result is not very encouraging, however, since other measurements don't violate any particular rule, reliability level can be considered acceptable.

After creating our scale, we compared the IIC of new scale and the original 22 items to see if our scale gives better results and suitable for factor score analysis. Figure 11 depicts the all item included model.

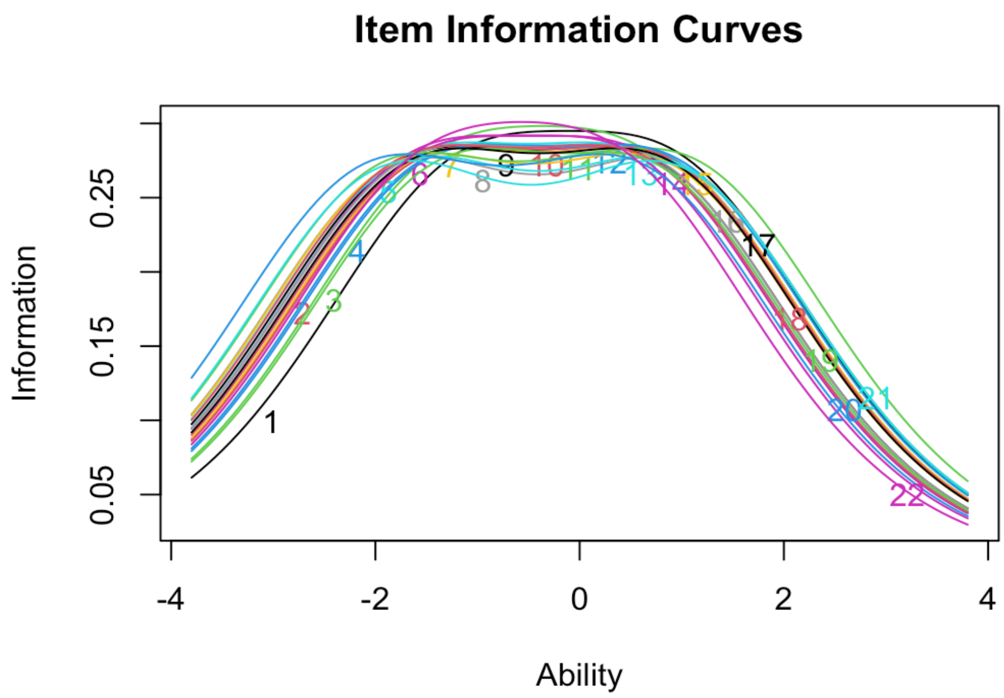


Figure 11. IIC for 22 items

It can be seen in Figure 11 for each item information levels and the ability of measuring the latent trait is roughly in the same level, since we used the constant discrimination model. Hence, we didn't compare the items, rather we checked the most information level, which in this case is quite low as approximately 0.27.

On the other hand, as shown in Figure 12, MSA applied constant discrimination GRM model's information levels peak around 0.41 which indicates a significant improvement.

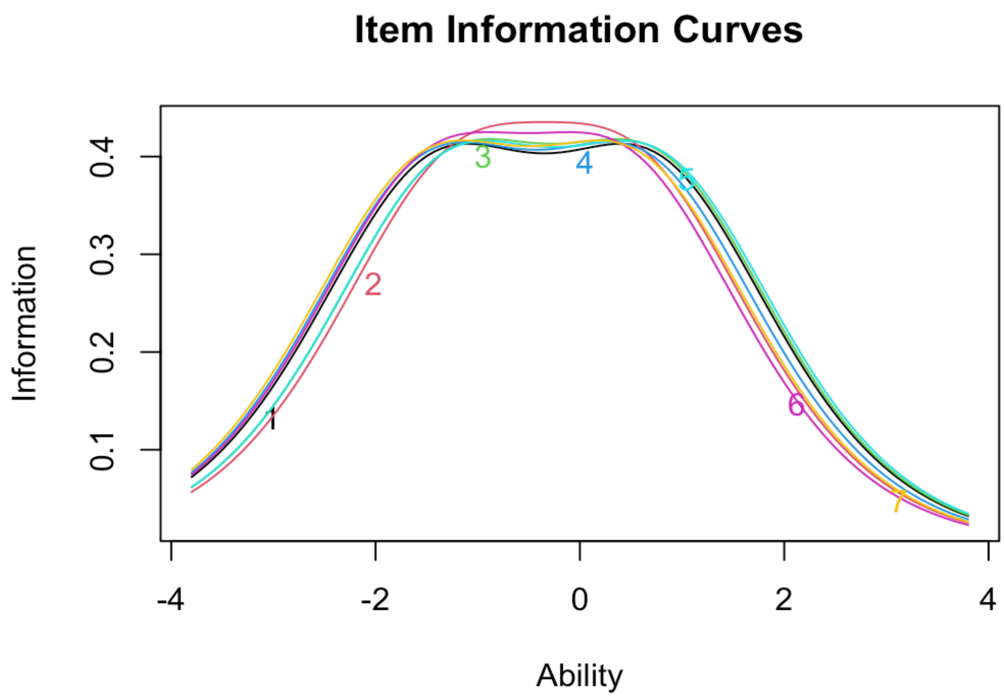


Figure 12. MSA applied constant discriminating GRM model

Additionally, we also created Test Information Curve for our model to see if there is any particular violation. From Figure 13, we can conclude that this test provides the most information in roughly -0.5 and 0.5 ability levels as expected.

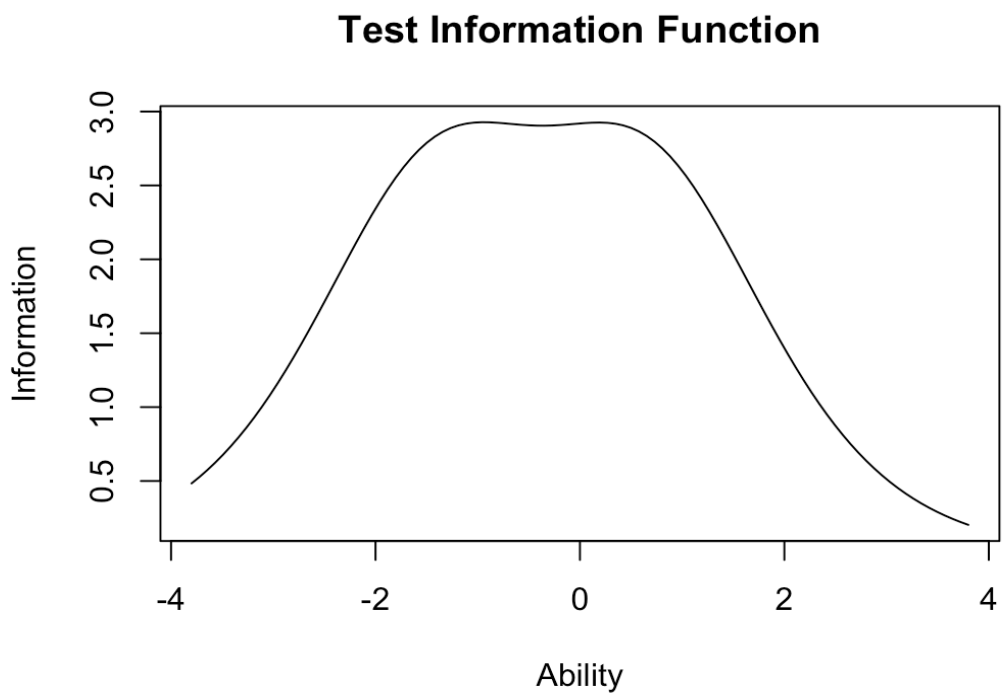


Figure 13. Test Information Curve

5.2.2 Factor Scores

Factor score analysis is pretty much the same procedure that is applied to Firm 1. Again, we are using *factor.score()* function in RStudio which performs CFA. Since the sample size is relatively higher than Firm 1, giving exact number of respondent opinion for each category is a little more challenging. Therefore, we created a scale that represents a category-like scoring. We took the average of each 212 response averages. The ones that are less than or equal to 1.5 consider their workplace as not stressful, scores between 1.5 and 2.5 ($1.5 < x \leq 2.5$) indicates they think the workplace or work-related issues are stressful and values greater than 2.5 means work related stress levels are critically high. According to this scale, 156 out of 212 people suffer from work related stress. Only 12 of them believe that there is no stress at work and 44 of them think they are having significantly high level of stress. In order to confirm the results and the effects over efficiency, factor scores are calculated for each response combination and mean value of these scores are found. Regarding the results, average factor score is roughly 0.744 for Firm 2. Finally, to validate the calculation of factor scores are properly measured by GRM. Kernel density estimation plot is created in Figure 14.

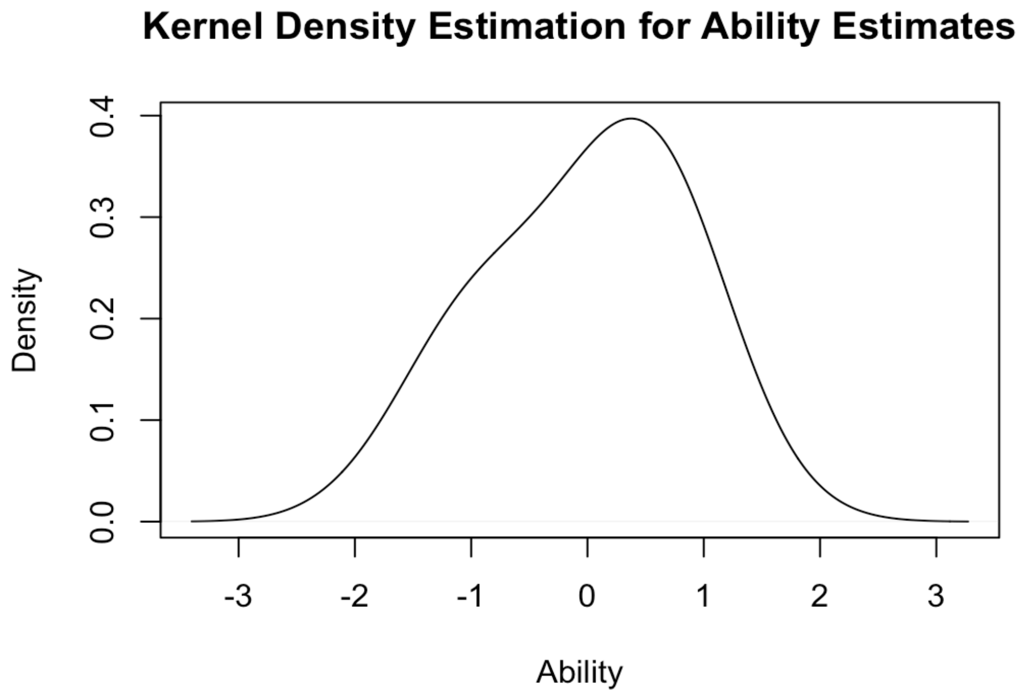


Figure 14. Kernel Density Estimation for GRM

Again, although in lower ability levels function slightly bends the form of normal distribution, it can be ignored since it is not a significant violation. Therefore, it can be concluded that GRM models factor scores are calculated successfully.

5.2.3 Technical Efficiency

TE calculation procedure is the same as Firm 1. Again, prior to calculating the efficiency levels, Shapiro-Wilk test is applied. The *p*-value has been estimated as 0.4511, which means efficiency levels are normally distributed. According to the procedures defined in Firm 1 TE analysis, output-based DEA efficiency results are given and reciprocal values are estimated in Table 24.

Table 24. Technical Efficiency Levels

	2020			
	September	**October**	**November**	**December**
Eff.	1.05	1.058	1.000	1.315
1/Eff.	0.952	0.944	1.000	0.969

	2021		
	January	**February**	**March**
Eff.	1.091	1.000	1.010
1/Eff.	0.916	1.000	0.989

Table 24 indicates that November and February are the most efficient months with 100 % efficiency and January is the least efficient one with roughly 91%.

5.2.4 Stress Effected Process

Before calculating SEP for Firm 2, TE with stress is calculated. Again, as in the previous section, factor scores are estimated in order to observe the stress pattern and its inevitable effect on productivity changes. Of course, before getting the results for SEP, Shapiro-Wilk test is applied to TE with stress model and resulted that the data is normally distributed with p-value of 0.421. Accordingly, efficiency levels are recalculated via output-based DEA model in Table 25.

Table 25. TE with stress

	2020			
	September	**October**	**November**	**December**
Eff.	1.078	1.176	1.000	1.032
1/Eff.	0.928	0.851	1.000	0.969

	2021		
	January	**February**	**March**
Eff.	1.122	1.055	1.031
1/Eff.	0.891	0.948	0.970

Note that these measures are sole measures for TE with stress. In order to find SEP, same procedure that is applied to Firm 1is used. Again, Equation (14) aided the SEP calculation and combined results are represented in Table 26.

Table 26. SEP for Firm 2

	September	October	November	December
SEP	0.97	0.90	1.00	1.00

	January	February	March
SEP	0.97	0.95	0.98

Unlike Firm 1, it is clear that there are several months, for example October, that might indicate a slightly stressful environment. However, before observing PC it is not proper to make a specific comment whether it is caused by SEP or TE.

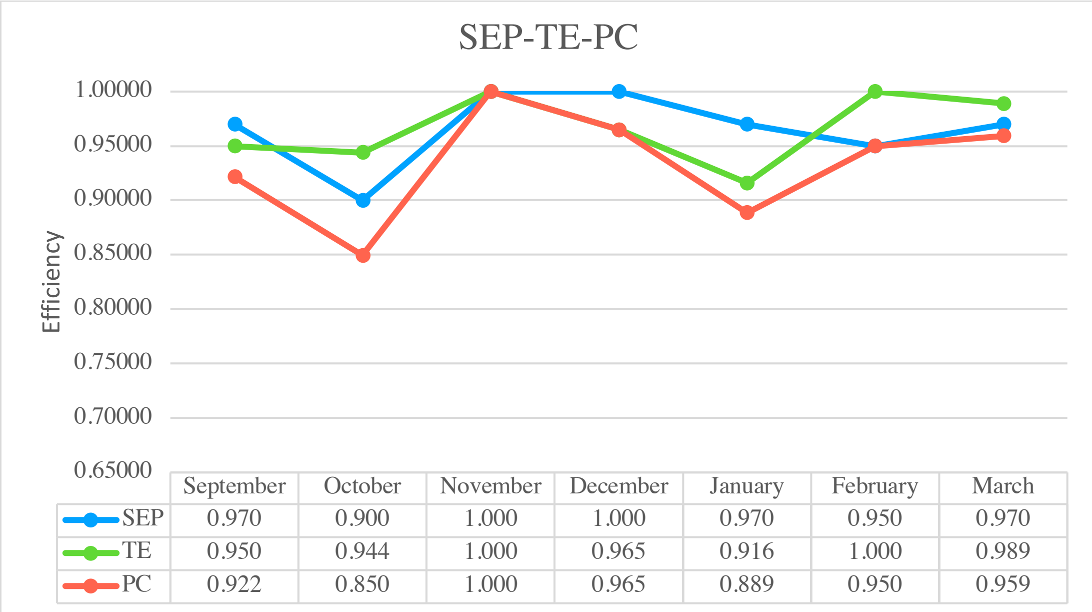
5.2.5 Productivity Changes

Concept has aided the study to depict the effects of both production efficiency and stress effects on firm overall productivity. Regarding the SEP and TE values for Firm 2, PC values are presented in Table 27.

Table 27. Combined values of SEP-TE and PC values

	SEP	TE	PC
September	0.970	0.950	0.922
October	0.900	0.944	0.850
November	1.000	1.000	1.000
December	1.000	0.965	0.965
January	0.970	0.916	0.889
February	0.950	1.000	0.950
March	0.970	0.989	0.959

It can be seen that both concepts (TE and SEP) have effects that can clearly be observed over PC for different months. For September, SEP has a higher value compared with TE. Hence, overall PC is mostly affected by the production process itself. On the other hand, in October, SEP has a significant effect over PC which means that in October, people tend to have more stress which affects their work efficiency and indirectly the overall firm productivity and efficiency. November is the optimal month according to SEP and TE values, and December is also optimal with being not stressful therefore PC values are solely affected by TE. Again, January's results are more affected by TE rather than SEP. Lastly, in February and March, work stress has more effect compared with TE to firm's overall productivity and efficiency. These results can be interpreted from Figure 15 as well.



	September	October	November	December	January	February	March
SEP	0.970	0.900	1.000	1.000	0.970	0.950	0.970
TE	0.950	0.944	1.000	0.965	0.916	1.000	0.989
PC	0.922	0.850	1.000	0.965	0.889	0.950	0.959

Figure 15. SEP-TE-PC graph

5.3 Item Analysis

In order to conclude our study, final examination is performed to determine what is perceived as the most stressful at work by the employees. This estimation is simply done by taking the average response rate for each item. For Firm 1, the rates are between 0 and 1 since we coded the questionnaire dichotomously. Higher rate means more stressful environment and vice versa. Accordingly, Table 28 is created.

Table 28. Average Response Rates for Firm 1

	Avg. R. Rate
Item 3	0.04
Item 4	0.14
Item 5b	0.52
Item 6b	0.32
Item 7b	0.22
Item 9b	0.48
Item 10b	0.26
Item 11b	0.24
Item 12b	0.4
Item13b	0.26
Item 14b	0.22
Item 16b	0.38
Item 17b	0.24
Item 18b	0.2
Item 21	0.34
Item 22b	0.46

Results shows that Items 5b and 9b have the highest rates with 0.5 and 0.48 respectively. This tells us people who work in Firm 1 think that workload increase and conflicts at work makes individuals more stressful compared with other elements. Items 3 and 4 have the least importance on estimating the stress with values of 0.04 and 0.14, which can be interpreted as people's opinions are moderately considered by their supervisor and workers have their will on deciding their work pace or have almost no complain about their pace. One item that we specifically want to highlight is Item 22b which is about current pandemic. It can be seen that people in Firm 1 don't find it as much stressful as it is imagined, however 0.46 is also one of the highest rates in this questionnaire; hence, it might be considered as well. These results are depicted in Figure 16 as well for ease of interpretation.

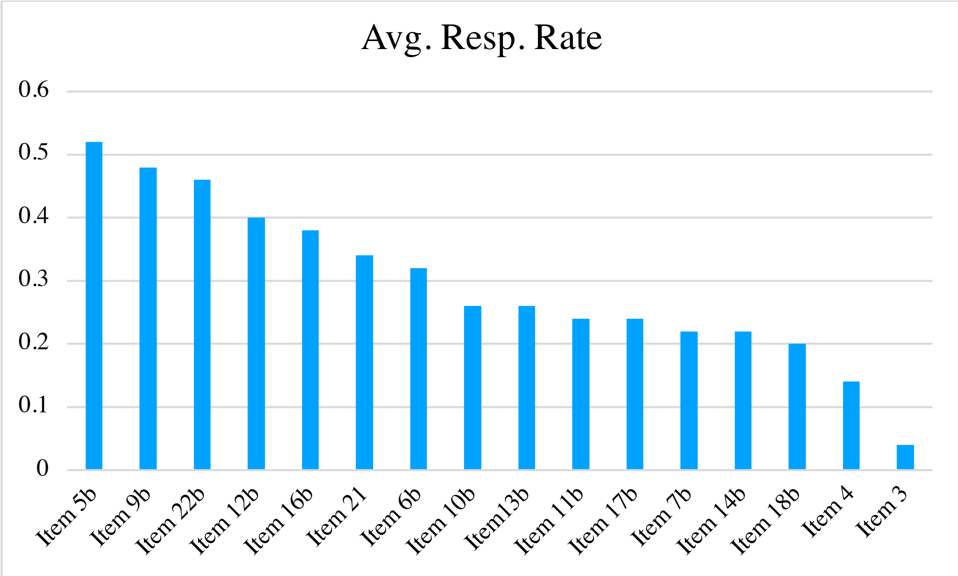


Figure 16. Average Response Rate's graph for Firm 1

Finally, same procedure is applied to Firm 2. This time the rates are between 1 and 3 since categories are polytomous. Table 29 displays average response rates for Firm 2.

Table 29. Average Response Rates for Firm 2

	Avg. R. Rate
Item 2	2.14
Item 3	2.16
Item 4	2.12
Item 9b	2.17
Item 12b	2.10
Item 14b	2.21
Item 18b	2.19
Item 22b	2.23

These values indicate that generally Firm 2 can be considered stressful since the values are greater than 1.5. Although we didn't include pandemic question in our scale, employees who work in this company think that pandemic is the most stressful element continuing with Item 14b with 2.21 rate, which tells us people tend to think about work even in their spare time which can be considered as a stressful phenomenon. These results can be read in Figure 17 as well.

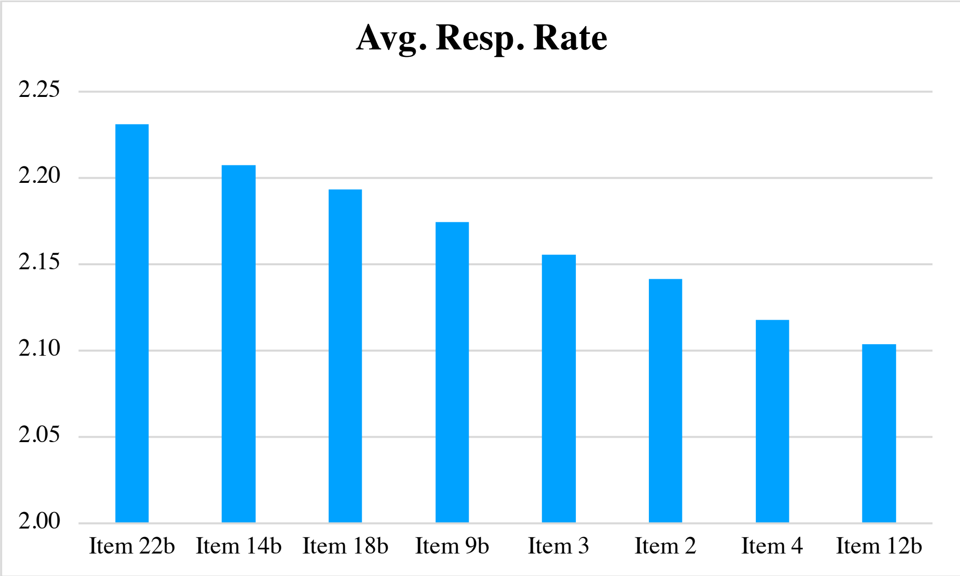


Figure 17. Average Response Rates for Firm 2

CHAPTER 6: CONCLUSION

The main goal of this thesis is to seek the potential indirect effects of stress to productivity and efficiency in 2 different manufacturing companies' plants in Turkey. Our study began with the collection of survey data and production data. In total, we roughly had 270 responses from both Firm 1's and Firm 2's employees. These results are prepared to get accurate results via dichotomous IRT and GRM. For production data, companies provided certain input and output values for us to calculate efficiency levels.

Our study proceeds with, dichotomous IRT and GRM models in order to determine the factor scores over Confirmatory Factor Analysis. For Firm 1, our sample size was significantly low so we preferably used dichotomous IRT to get proper results. Fortunately, Firm 2's participation rate was suitable for GRM model.

We continued our study with developing a Malmquist-like output-based DEA model to calculate *TE, SEP* and *PC* values for each firm individually to interpret the different efficiency values for specified input and output values. Results are discussed diligently for each firm. We finalized our research with Item Analysis to understand the stress pattern for each firm.

Results indicate that Firm 1 doesn't seem to be a stressful company as their SEP is almost always optimal during 7 months process. On the other hand, we have substantial evidence that employees who work at Firm 2 might be overwhelmed by stress in months like October and March which causes company to lack efficiency and productivity. Additionally, item analysis tells that workload increase, conflict at work and focusing on work in spare times might have more impact on increasing the stress levels. This justifies that our study also provides potential causes of stress which creates an important area to work on.

We think these results can be used to enhance worker's mental health in order to fulfill the study purpose. Further, it is possible to explore the months that has higher level of stress to understand the stress pattern and their relation with certain features like seasonality. Study also aids further studies about industry dependency of work stress as well. We believe that this is an area which is substantial for both worker's health as well as companies' mission.

REFERENCES

Afriat, S.N. (1972) *Efficiency Estimation of Production Function,* International Economic Review, Vol.13(3), pp. 568-598.

Ali, A.I., Lerme, C. (1997) *Comparative advantage and disadvantage in DEA*, Annals of Operations Research, Vol.73(0), pp. 215-232.

Armagan, G., Ozden, A. and Bekcioglu, S. (2010) *Efficiency and total factor productivity of crop production at NUTS1 level in Turkey: Malmquist index approach.*, Quality & Quantity, Vol.44, pp. 573–581.

Berg, S. A., Forsund, F.R. and Jansen, E.S. (1992) *Malmquist indexes of productivity growth during the deregulation of Norwegian banking*, Scandinavian Journal of Economics, Vol.94, pp. 211-228.

Berg, S.A., Forsund, F., Hjalmarsson, L. and Suominen, M. (1993) *Banking efficiency in the Nordic countries*, Journal of Banking & Finance, Vol.17(2-3), pp. 371-388.

Berg, S.A., Forsund, F.R. and Jansen, E.S. (1992) *Malmquist Indices of Productivity Growth during the Deregulation of Norwegian Banking,1980-89,* Scandinavian Journal of Economics, Vol.94, pp.211-228.

Bessent, A., Bessent E., Kennington J. and Reagan, B. (1980) *An application of mathematical-programming to assess productivity in the Houston independent school-district*, Management Science, Vol.28(12), pp. 1355-1475.

Black C. (2008) *Working for a Healthier Tomorrow-Dame Carol Black's Review of the Health of the Working Age Population*. Presented to the Secretary of State for Health and the Secretary of State for Work and Pensions, London: TSO.

Blumenfeld, D. and Inman, R. (2009) *Impact of absenteeism on assembly line quality and throughput,* Production and Operations Management, Vol. 18 (3), pp. 333-343.

Brown, T. A., and Moore, M. T. (2012) *Confirmatory factor analysis*, In R. H. Hoyle (Ed.), *Handbook of structural equation modeling,* (pp. 361–379). The Guilford Press.

Caves, D.W., Christensen, L.R. and Diewert, W.E. (1982) *The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity*, Econometrica, Vol.50(6), pp. 1393-1414.

Charles A., Cooper W.W. and Rhodes E. (1978) *Measuring the Efficiency of Decision-Making Units*, European Journal of Operations Research, Vol. 2(6), pp. 429 – 444.

Chavas, J.P., Aliber, M. (1993) *An analysis of economic efficiency in agriculture—a nonparametric approach*, Journal of Agricultural and Resource Economics, Vol. 18(1), pp. 1-16.

Choi, J. and Roberts, D. (2015) *How Does the Change of Carbon Dioxide Emissions Affect Transportation Productivity? A Case Study of the US Transportation*, Open Journal of Social Sciences, Vol.3, pp. 96-106.

UCLA: Statistical Consulting Group (2021),
Confirmatory Factor Analysis (CFA) in R with lavaan, [Online], Available at: https://stats.idre.ucla.edu/r/seminars/rcfa/ (Accessed: 26 Oct. 2021).

Cramér, H. (1946) Mathematical *methods of statistics (9th edition),* New York: Princeton University Press.

Davidson, R. And Flachaire, E. (2008) *The wild bootstrap, tamed at last,* Journal of Econometrics, Vol.146(1), pp. 162-169.

Depaoli, S., Tiemensma, J. and Felt, J.M. (2018) *Assessment of health surveys: fitting a multidimensional graded response model,* Psychology, Health & Medicine, Vol.23(1), pp. 1299-1317.

Business News Daily. (2020) *Do you see signs of workplace stress at your company?* [Online] Available at: https://www.businessnewsdaily.com/2267-workplace-stress-health-epidemic-perventable-employee-assistance-programs.html (Accessed: 3 Oct. 2021).

Ebert, T. (2018), *Re: What are some of the disadvantages of data bootstrapping?* [Online] Available at: https://www.researchgate.net/post/What_are_some_of_the_disadvantages_of_data_bootstrapping. (Accessed: 3 Oct. 2021).

Efron, B. (1979) *Bootstrap methods: another look at the jackknife*, The Annals of Statistics, Vol.7(1), pp. 1-26.

Embretson, S. E., & Reise, S. P. (2000) *Item Response Theory for Psychologists (1st Edition),* New Jersey: Psychology Press

Epstein, M.K. and Henderson, J.C. (1989) *Data envelopment Analysis for managerial control and diagnosis,* Decision Sciences, Vol.20(1), pp.90-119.

Färe, R., Grabowski, R. and Grosskopf, S. (1985) *Technical efficiency of Philippine agriculture*, Applied Economics, Vol.17(2), pp. 205-214.

Färe, R., Grosskopf, S. and Norris, M. (1997) *Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countrie*s: *Reply,* The American Economic Review, Vol. 87(5), pp. 1040-1044.

Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994) *Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countrie*s, The American Economic Review, Vol. 84(1), pp. 66-83.

Farrell, M.J. (1957) *The Measurement of Productive Efficiency,* Journal of The Royal Statistical Society, Vol. 120(3), pp. 253-281.

Gambhir, D. and Sharma, S. (2015) *Productivity in Indian manufacturing: evidence from the textile industry*, Journal of economics and Administrative Sciences, Vol.31(2), pp.71-85.

Grifell-Tatje, E. and Lovell, C.A.K. (1995) *A note on the Malmquist productivity index*, Economics Letters, Vol.47(2), pp. 169-175.

Gummelt, H. D., Anestis, J. C., & Carbonell, J. L. (2012). *Examining the Levenson Self Report Psychopathy Scale using a Graded Response Model*. Personality and Individual Differences, Vol.58(8), pp.1002-1006.

Harter, J. (2017) *Dismal Employee Engagement Is Sign of Global Mismanagement*, [Online] Available at: https://www.gallup.com/workplace/231668/dismal-employee-engagement-sign-global-mismanagement.aspx (Accessed: 3 Oct. 2021).

Holmgren, K. and Frantz, A. (2019) *The Work Stress Questionnaire (WSQ) – reliability and face validity among male workers*, BMC Public Health, Vol.19, Available at: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-019-7940-5#citeas (Accessed: 25 Feb. 2022).

Holmgren, K., Dahlin Ivanoff, S., Bjorkelund, C. and Hensing, G. (2009) *The prevalence of work-related stress, and its association with self-perceived health and sick-leave, in a population of employed Swedish women*. BMC Public Health, Vol.9(1) Available at: https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-9-73 (Accessed: 25 Feb. 2022).

Holmgren, K., Fjallstrom-Lundgren, M., Hensing, G. (2013) *Early identification of work-related stress predicted sickness absence in employed women with musculoskeletal or mental disorders. A prospective, longitudinal study in a primary health care setting*, Disability and Rehabilitation, Vol.35(5), pp. 418-426.

Holmgren, K., Hensing, G. and Dahlin-Ivanoff, S. (2009) *Development of a questionnaire assessing work-related stress in women - identifying individuals who risk being put on sick leave,* Disability and Rehabilitation, Vol.31(4), pp. 284-292.

Hudaefi, F., Caraka, R. and Wahid, H. (2021) *Zakat Administration in Times of COVID-19 Pandemic in Indonesia: A Knowledge Discovery via Text Mining*. International Journal of Islamic and Middle Eastern Finance and Management, Available at: https://www.researchgate.net/publication/350891868_Zakat_Administration_in_Times_of_COVID-

19_Pandemic_in_Indonesia_A_Knowledge_Discovery_via_Text_Mining (Accessed: 25 Feb. 2022).

Javelle, F., Wiegand, M., Joormann, J., Timpano, K. R., Zimmer, P. and Johnson, S. L. (2021) *The German Three Factor Impulsivity Index: Confirmatory factor analysis and ties to demographic and health-related variables,* Personality and Individual Differences, Vol.171, Available at: https://pubmed.ncbi.nlm.nih.gov/35185234/ (Accessed: 25 Feb. 2022)

Jiang, S., Wang, C. and Weiss, D. (2016) *Sample Size Requirements for Estimation of Item Parameters in the Multidimensional Graded Response Model*, Frontiers in Psychology, Vol.7, Available at : https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00109/full (Accessed: 25 Feb. 2022)

Johansen, L. (1968) *Production Functions and the Concept of Capacity*, Centre d'Etudes et de Recherches Universitaire de Namur (Ceruna), Collection Economie Mathématique et Econométrie, Vol.2, pp. 49–72.

Jöreskog, K.G. (1969) *A General Approach to Confirmatory Factor Analysis,* Psychometrika, Vol.34(2), pp. 183-202.

Lazarsfeld, P.F. (1950) *The logical and mathematical foundation of latent structure analysis*. In: Stouffer, S.A., Guttman, L., Suchman, E.A., Lazarsfeld, P.F., Star, S.A., Clausen, J.A. (Eds.), Measurement and Prediction. Wiley, New York, pp. 362–412.

Le Cam, L. "*On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates*", (1960) Matematika, Vol.4(2), pp. 69–120; (1953) *Univ. Calif. Public. Statist*. Vol.1(11), pp. 277-330.

Liu, S. Lu, L., Lu, W. and Lin, B. (2013), *The survey of DEA applications*, Omega, Vol.41(5), pp. 893-902.

Matthews G., *Distress* (2000) In: Encyclopedia of Stress, Vol.1, San Diego, CA: Academic Press.

McMullen, B.S. and Okuyama, K. (1996) *Productivity Changes in the U.S. Motor Carrier Industry 1976-1990: A Malmquist Index Approach,* Transportation Research Forum, 38th Annual Meeting, Volume 2, pp.516-532.

Meijer, R. R. and Baneke, J. J. (2004*) Analyzing Psychopathology Items: A Case for Nonparametric Item Response Theory Modeling,* Psychological Methods, Vol.9(3), pp.354–368.

Michie, S. (2002) *Causes and Management of Stress at Work,* Occupational and Environmental Medicine, Vol.59(1), pp. 67-72.

Molenaar, I.W. (1997) *Nonparametric Models for Polytomous Responses*. In: Van der Linden W.J., Hambleton R.K. (eds) Handbook of Modern Item Response Theory. New York: Springer.

Myung, J. (2003) *Tutorial on maximum likelihood estimation,* Journal of Mathematical Psychology, Vol.47, pp.90-100.

Neyman, J. and Scott, E.L. (1948) *Consistent Estimates Based on Partially Consistent Observations*, Econometrica, Vol.16(1), pp. 1-32.

Norden, R.H. (1972) *A Survey of Maximum Likelihood Estimation*, International Statistical Review, Vol.40(3), p.329-354.

Norsworthy, J.R., Malmquist, D. (1983) *Input measurement and productivity growth in Japanese and US manufacturing,* American Economic Review, Vol.73(5), pp.1020-1033.

Nunamaker, TR. (1983) *Measuring routine nursing service efficiency: a comparison of cost per patient day and data envelopment analysis models*. Health Services Research; Vol.18(2 Pt 1) pp. 183-208.

Ødegaard, F. and Roos, P. (2014) *the Contribution of Workers' Health and Psychosocial Work-Environment on Production Efficiency*, Production and Operations Management, Vol. 23(12), pp. 2191-2208.

Qazi, A.Q. and Yulin, Z. (2012) *Productivity Measurement of Hi-tech Industry of China Malmquist Productivity Index – DEA Approach*, Procedia Economics and Finance, Vol.1, pp. 330-336.

Rajasekar, T., Deo*, M*. (2014) *Is There Any Efficiency Difference between Input and Output Oriented DEA Models: An Approach to Major Ports in India*, Journal of Business & Economic Policy, Vol.1(2). Available at: https://jbepnet.com/journals/Vol_1_No_2_December_2014/ 2.pdf (Accessed: 25 Feb. 2022).

Rangan, N., Grabowski, R., Aly, H. and Pasurka, C. (1988) *The technical efficiency of United States banks*, Economics Letters, Vol. 28(2), pp. 169-175.

Rao, C.R. (1960) *Apparent anomalies and irregularities of the method of maximum likelihood*, Proceedings of the 32nd session of the Int. Stat. Inst., Tokyo.

Reeve, B. (2002) *An introduction to modern measurement theory*. Bethesda: National Cancer Inst. Available at: http://coshima.davidrjfikis.com/EPRS8410/Angela_EPRS9350_IRT_NCI_Reeves_Tutorial.pdf (Accessed: 25 Feb. 2022).

Reise, S.P. and Waller, N.G. (2003) *How many IRT parameters does it take to model psychopathology items?* Psychological Methods, Vol.8(2), pp. 164-184.

Reise, S.P., Ainsworth, A.T. and Haviland, M.G. (2005) *Item Response Theory: Fundamentals, Applications, and Promise in Psychological Research*, Sage Journals, Vol.14(2), pp. 95-101.

Samejima F. (1997) *Graded Response Model*. In: van der Linden W.J., Hambleton R.K. (eds*) Handbook of Modern Item Response Theory*. New York: Springer

Samejima, F. (1969), *Estimation of latent ability using a response pattern of graded scores*, Psychometrika, Vol.34, pp. 1-97.

Schefczyk*, M.,* (1993) *Operational performance of airlines—an extension of traditional measurement paradigms*, Strategic Management Journal, Vol.14, pp. 301-317.*Sector from 2002 to 2011*. Open Journal of Social Sciences, Vol.3, pp. 96-106.

Schwartz, S.H. and Boehnke, K. (2004), *Evaluating the structure of human values with confirmatory factor analysis*, Journal of Research in Personality, Vol.38(3), pp.230-255

Sherman, D., Gold, F., (1985) *Bank branch operating efficiency—evaluation with data envelopment analysis*. Journal of Banking & Finance, Vol. 9(2), pp. 297-315.

Sherman, HD., (1984) *Hospital efficiency measurement and evaluation. Empirical test of a new technique*. Medical Care, Vol.22(10), pp. 922-38.

Siengthai, S., Sukirno, D. (2010) *The Comparison of Graded Response Model and Classical Test Theory in Human Resource Research: A Model Fitness Test*, Research and Practice in Human Resource Management Journal, Vol.18(2), pp.77-90.

Sijtsma, K. and Molenaar, IW., (2002), *Introduction to nonparametric item response theory*. Measurement methods for the Social Science, no. 5, California: Sage

Singh, K. (1981) *On the asymptotic accuracy of Efron's bootstrap,* The Annals of Statistics, Vol.9(6), pp.1187-1195.

Sowlati, T. and Vahid, S. (2006) *Malmquist productivity index of the manufacturing sector in Canada from 1994 to 2002, with a focus on the wood manufacturing sector,* Scandinavian Journal of Forest Research, Vol.21(5), pp.424-433.

Stănculescu E. (2021) *Fear of COVID-19 in Romania: Validation of the Romanian Version of the Fear of COVID-19 Scale Using Graded Response Model Analysis*, International Journal of Mental Health Addiction. Vol.6, pp. 1-16

Suhr, D. (2006), *Exploratory or Confirmatory Factor Analysis?* Proceedings of the 31st Annual SAS. Users Group International Conference. Cary, NC: SAS Institute Inc. pp. 200-231.

Taskin, F. and Zaim, O. (2000*) Searching for a Kuznets curve in environmental efficiency using kernel estimation,* Economics Letters, Vol.68(2), pp. 217-223.

Tauer, L. (1994) *Efficiency, Technology, and Productivity Changes on Individual Dairy Farms*, Working Papers 179226, Cornell University, Department of Applied Economics and Management.

Thirtle, C., Hadley, D. and Townsend R. (1994) *Policy Induced In- novation in Sub-Saharan African Agriculture: A Multilateral Malmquist Productivity Index Approach*, Development Policy Review, Vol.13(4), pp. 323-348.

Thomas, M.L. (2011) *The value of item response theory in clinical assessment: a review,* Assessment, Vol.18(3), pp. 291-307.

Thurstone, L.L. (1947) *Multiple Factor Analysis,* Chicago: University of Chicago Press.

Tone, K. (2006) M*almquist Productivity Index: Efficiency Change Over Time,* Handbook on Data Envelopment Analysis*,* Chapter 8, pp. 203-227.

Tyrone T. Lin, Lee, C.-C. and Chiu, T.-F (2009) *Application of DEA in analyzing a bank's operating performance*, Expert Systems with Applications, Vol.36(5), pp. 8883-8891.

Van Dam, N.T., Earleywine, M. and Borders, A. (2010) *Measuring mindfulness? An Item Response Theory analysis of the Mindful Attention Awareness Scale*, Personality and Individual Differences, Vol.49(7), pp. 805-810.

Van der Ark, L.A. (2007) *Mokken Scale Analysis in R*, Journal of Statistical Software, Vol.20(11), pp. 1–19

Van der Ark, L.A. (2012) *New Developments in Mokken Scale Analysis in R*, Journal of Statistical Software, Vol.48(5), pp.1–27.

Van der Linden, W.J. (2005) *Item Response Theory*, Encyclopedia of Social Measurement, pp. 379-387.

Wadud, A. and White, B. (2000) *Farm household efficiency in Bangladesh: A comparison of stochastic frontier and DEA methods,* Journal of Applied Economics, Vol. 32, Issue13, pp. 1666 -1673.

Ward, P.T., Storbeck, J.E., Mangum, S.L. and Byrnes, P.E. (1997) *The analysis of staffing efficiency in U.S. manufacturing:1983 and 1989*, Annals of Operations Research, Vol.73, pp.67-89.

Wilson, P. (2008) *FEAR*: *A software package for frontier efficiency analysis with R,* Journal of Socio-Economic Planning Sciences, Vol.42(4), pp.247-254.

Zhang, H. & Wang, X.-B. (2005) *Green manufacturing process assessment by DEA method*, Binggong Xuebao/Acta Armamentarii, Vol.26. pp.523-527.

APPENDICES

Appendix A- Work Stress Questionnaire

The Work Stress Questionnaire (revised version)

01 Do you have time to finish your assignments?	□ yes, always □ yes, rather often □ no, seldom □ no, never
02 Do you have the possibility to influence decisions at work?	□ yes, always □ yes, rather often □ no, seldom □ no, never
03 Does you supervisor consider your views?	□ yes, always □ yes, rather often □ no, seldom □ no, never
04 Can you decide on your work pace?	□ yes, always □ yes, rather often □ no, seldom □ no, never
05a Has your workload increased?	□ yes □ no – if no: go to question **06a**
05b If *yes*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
06a Are the goals for your workplace clear?	□ yes – if *yes* continue to question **07a** □ partly □ no
06b If *partly* or *no*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
07a Do you know which assignments your work tasks include?	□ yes – if *yes* continue to question **08a** □ partly □ no
07b If *partly* or *no*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
08a Do you know who is making decisions concerning your workplace?	□ yes – if *yes* continue to question **09a** □ partly □ no
08b If *partly* or *no*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
09a Are there any conflicts at work?	□ yes □ no – if *no* continue to question **12a**
09b If *yes*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
10a Are you involved in any conflicts at your workplace?	□ yes □ no – if *no* continue to question **11a**
10b If *yes*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
11a Have your supervisor done anything to solve the conflicts?	□ yes – if *yes* continue to question **12a** □ partly □ no
11b If *partly* or *no*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful

12a Do you put high demands on yourself at work?	□ yes □ no – if *no* continue to question **13a**
12b If *yes*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
13a Do you often get engaged in your work?	□ yes □ no – if *no* continue to question **14a**
13b If *yes*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
14a Do you think about work after your working-day?	□ yes □ partly □ no – if *no* continue to question **15a**
14b If *yes* or *partly*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
15a Do you find it hard to set a limit to work assignment although you have a lot to do?	□ yes □ partly □ no – if *no* continue to question **16a**
15b If *yes* or *partly*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
16a Do you take more responsibility at work than you ought to?	□ yes □ no – if *no* continue to question **17a**
16b If *yes*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
17a Do you work after ordinary working hours to finish your assignments?	□ yes □ partly □ no – if *no* continue to question **18a**
17a If *yes* or *partly*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
18a Do you find it hard to sleep because your mind is occupied with work?	□ yes □ partly □ no – if *no* continue to question **19**
18b If *yes* or *partly*: Do you perceive that as stressful?	□ not stressful □ less stressful □ stressful □ very stressful
19 Due to work, do you find it hard to find time to be with your nearest?	□ yes, always □ yes, rather often □ no, seldom □ no, never
20 Due to work, do you find it hard to find time to be with your friends?	□ yes, always □ yes, rather often □ no, seldom □ no, never
21 Due to work, do you find it hard to find time for your recreational activities?	□ yes, always □ yes, rather often □ no, seldom □ no, never

Appendix B- WSQ Turkish

İŞ STRES ANKETİ

BİLGİLENDİRME (LÜTFEN DİKKATLİCE OKUYUNUZ!)

1. Bu anket firma çalışanlarının işten kaynaklanan stresini ölçmek üzere tasarlanmıştır.
2. Anketin orijinal adı Work Stress Questionnaire (WSQ) olup, İsveç'li bilim insanları tarafından geliştirilmiştir
3. Anket uzun yıllar içerisinde geliştirilmiş olup sürekli güncellenmektedir. Bu bağlamda güvenilir bir kaynak oluşturmaktadır.
4. Ankette sağlayacağınız kişisel veriler (yaş ve cinsiyet) yalnızca araştırmacı tarafından görülebilecek ve hiç bir koşulda kimse ile bireysel anket olarak paylaşılmayacaktır.
5. Anket İzmir Ekonomi Üniversitesi Etik değerlendirme kurulu tarafından değerlendirilmiş ve onaylanmıştır.
6. Anket güvenilirliği ve verilerin doğru sonuç vermesi için lütfen anket sorularını iyi anladığınıza ve dürüstçe yanıtladığınıza emin olunuz.
7. Anket süresince bütün soruları sırayla yanıtlamanız gerekmektedir. Ancak anketteki bazı sorular verdiğiniz yanıta göre cevaplanmak zorunda değildir. Böyle bir durumda anket içindeki yönergeyi takip ediniz.
8. Anketi tamamladığınızda anket dağıtıcısının size söylediği anket toplama kutusuna anketi bırakmalısınız.
9. Çalışmayla ilgili her türlü sorunuzu ve öğrenmek istediklerinizi ozturkonurarin@gmail.com adresine mail atabilirsiniz.

Anketi Yapacak kişinin ;

Yaşı :
Cinsiyeti :
Eğitim Durumu :

Sorular

1. Verilen işi bitirmek için yeterli zamanınız var mı ?

O Evet, her zaman	O Evet, çoğunlukla	O Hayır, nadiren	O Hayır, hiçbir zaman

2. İş yerinde kararları etkileme olanağınız var mı ?

O Evet, her zaman	O Evet, çoğunlukla	O Hayır, nadiren	O Hayır, hiçbir zaman

3. Amiriniz (müdürünüz, bağlı olduğunuz departman sorumlusu vb.) görüşlerinizi değerlendiriyor mu?

O Evet, her zaman	O Evet, çoğunlukla	O Hayır, nadiren	O Hayır, hiçbir zaman

4. Çalışma hızınıza karar verebiliyor musunuz?

O Evet, her zaman	O Evet, çoğunlukla	O Hayır, nadiren	O Hayır, hiçbir zaman

5a. İş yükünüz çalıştığınız süre boyunca hiç arttı mı?

O Evet	O Hayır

(Eğer Cevabınız 'Hayır' ise soru 6a'ya geçin)

5b. Eğer cevabınız 'Evet' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

6a. İş yerinizin hedefleri anlaşılır ve net mi?

O Evet	O Kısmen	O Hayır

(Eğer cevabınız Evet ise soru 7a'ya geçin)

6b. Eğer cevabınız 'Kısmen' veya 'Hayır' ise: Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

7a. İşinizin hangi görevleri içerdiğini net bir şekilde biliyor musunuz?

O Evet	O Kısmen	O Hayır

(Eğer cevabınız Evet ise soru 8a'ya geçin)

7b. Eğer cevabınız 'Kısmen' veya 'Hayır' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

8a. İş yeri ile ilgili kararları kimlerin verdiğini biliyor musunuz?

O Evet	O Kısmen	O Hayır

(Eğer cevabınız Evet ise soru 9a'ya geçin)

8b. Eğer cevabınız 'Kısmen' veya 'Hayır' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

9a. İş yerinizde herhangi bir çatışma var mı? (Fikir çatışması vb.)

O Evet	O Hayır

(Eğer cevabınız Hayır ise soru 12a'ya geçin)

9b. Eğer cevabınız 'Evet' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

10a. İş yerinde herhangi bir çatışmaya dahil oldunuz mu?

O Evet	O Hayır

(Eğer cevabınız 'Hayır' ise soru 11a'ya geçin)

10b. Eğer cevabınız 'Evet' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

11a. Amirleriniz iş yerindeki çatışmaları çözmek için herhangi bir şey yapıyor mu ?

O Evet	O Kısmen	O Hayır

(Eğer cevabınız Evet ise soru 12a'ya geçin)

11b. Eğer cevabınız 'Kısmen' veya 'Hayır' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

12a. İş yerinizde kendinize fazla yüklendiğinizi hissediyor musunuz?

O Evet	O Hayır

(Eğer cevabınız Hayır ise soru 13a'ya geçin)

12b. Eğer cevabınız 'Evet' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

13a. İşinizle sık sık meşgul oluyor musunuz?

O Evet	O Hayır

(Eğer cevabınız Hayır ise soru 14a'ya geçiniz)

13b. Eğer cevabınız 'Evet' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

14a. Mesai dışında işinizi düşünüyor musunuz?

O Evet	O Kısmen	O Hayır

(Eğer cevabınız 'Hayır' ise soru 15a' ya geçiniz)

14b. Eğer cevabınız 'Evet' veya 'Kısmen' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

15a. Yapacak çok işiniz olmasına rağmen iş ataması için bir sınır belirlemekte zorlanıyor musunuz?

O Evet	O Kısmen	O Hayır

(Eğer cevabınız 'Hayır' ise soru 16a'ya geçin)

15b. Eğer cevabınız 'Evet' veya 'Kısmen' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

16a. İş yerinde yapmanız gerekenden daha fazla sorumluluk alıyor musunuz?

O Evet	O Hayır

(Eğer cevabınız 'Hayır' ise soru 17a'ya geçin)

16b. Eğer cevabınız 'Evet' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

17a. Normal mesai saatleriniz dışında görevlerinizi bitirmek için fazla mesai yapıyor musunuz?

O Evet	O Kısmen	O Hayır

(Eğer cevabınız 'Hayır' ise soru 18a'ya geçiniz)

17b. Eğer cevabınız 'Evet' veya 'Kısmen' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

18a. İş ile ilgili konuları düşünmekten kaynaklanan uyku problemi yaşıyor musunuz?

O Evet	O Kısmen	O Hayır

(Eğer cevabınız 'Hayır' ise soru 19'a geçiniz)

18b. Eğer cevabınız 'Evet' veya 'Kısmen' ise : Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

19. İşinizden dolayı ailenize vakit ayırmakta zorlanıyor musunuz?

O Evet, her zaman	O Evet, çoğunlukla	O Hayır, nadiren	O Hayır, hiçbir zaman

20. İşinizden dolayı arkadaşlarınıza vakit ayırmakta zorlanıyor musunuz?

O Evet, her zaman	O Evet, çoğunlukla	O Hayır, nadiren	O Hayır, hiçbir zaman

21. İşinizden dolayı eğlence veya dinlenme amaçlı aktivitelere zaman ayırmakta zorlanıyor musunuz?

O Evet, her zaman	O Evet, çoğunlukla	O Hayır, nadiren	O Hayır, hiçbir zaman

22a. Covid-19 pandemisinin işinizi etkilediğini düşünüyor musunuz?

O Evet	O Kısmen	O Hayır

(Eğer cevabınız 'Hayır' ise anketi bitiriniz)

22b. Eğer cevabınız 'Evet' veya 'Kısmen' ise :Bu durumu stresli buluyor musunuz?

O Hiç stresli değil	O Az stresli	O Stresli	O Çok stresli

Ankete katıldığınız için teşekkürler.

Appendix C- Firm 1 Factor Score Table

Item 3	Item 4	Item 5b	Item 6b	Item 7b	Item 9b	Item 10b	Item 11b	Item 12b	Item 13b	Item 14b	Item 16b	Item17b	Item 18b	Item 21	Item 22b	Obs.	Exp.	z1	se.z1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	5.9730	0.0000	0.6172
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0.1848	0.0000	0.4442
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0.0108	0.0000	0.3824
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0.3884	0.0000	0.5215
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0.1130	0.0000	0.4492
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0.2056	0.0000	0.4847
0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	0.0275	0.0017	0.3467
0	0	0	0	0	1	0	0	0	0	1	0	1	1	0	0	1	0.0005	0.1285	0.3306
0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	2	0.1040	0.0000	0.4283
0	0	0	0	1	1	0	0	1	0	0	1	0	0	0	1	1	0.0057	0.4174	0.3060
0	0	0	0	1	1	0	1	1	0	1	1	1	1	1	1	1	0.0009	1.1541	0.3095
0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	1	0.0028	0.0760	0.3369
0	0	0	1	0	1	1	0	0	0	0	0	0	1	0	0	1	0.0020	0.0241	0.3436
0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	1	0.0012	0.0724	0.3373
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.7571	0.0000	0.5211
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0.5121	0.0000	0.4490
0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0.0378	0.0000	0.4050
0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0.1892	0.0000	0.4129
0	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	1	0.0089	0.0349	0.3422
0	0	1	0	0	0	0	0	1	1	1	1	1	0	1	1	1	0.0012	0.7444	0.2966
0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0.0063	0.0695	0.3377
0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0.0530	0.0904	0.3351
0	0	1	0	0	1	0	1	1	0	0	1	0	1	1	1	1	0.0045	0.7390	0.2966
0	0	1	0	0	1	0	1	1	0	0	1	1	1	0	1	1	0.0024	0.7292	0.2966
0	0	1	0	0	1	1	0	0	0	0	1	0	0	1	1	1	0.0092	0.3346	0.3114
0	0	1	0	0	1	1	1	1	0	0	0	0	0	0	1	1	0.0053	0.3315	0.3116
0	0	1	0	0	1	1	1	1	1	0	1	0	0	1	0	1	0.0022	0.7192	0.2967
0	0	1	0	1	0	0	0	0	1	1	1	0	0	1	1	1	0.0006	0.6488	0.2975
0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0.0784	0.0000	0.3792
0	0	1	1	0	1	0	0	0	1	0	1	0	0	1	1	1	0.0062	0.5309	0.3006
0	0	1	1	0	1	1	0	1	0	0	0	0	0	1	0	1	0.0049	0.3182	0.3126
0	0	1	1	0	1	1	1	1	0	0	0	0	0	0	0	1	0.0025	0.3539	0.3100
0	0	1	1	1	1	0	0	1	0	0	0	0	0	1	1	1	0.0047	0.5391	0.3003
0	0	1	1	1	1	1	1	0	0	0	1	1	0	0	0	1	0.0004	0.6945	0.2969
0	1	1	0	0	0	0	0	1	1	1	1	1	0	0	0	1	0.0002	0.6997	0.2968
0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	1	0.0032	0.1929	0.3237
0	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	0.0084	1.5393	0.3456
0	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	0.0013	1.3795	0.3279
0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0.0007	1.2263	0.3145
0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0.0908	1.9070	0.3987
1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.0026	1.8598	0.3910
1	1	0	1	1	0	0	0	0	0	0	1	1	0	0	1	1	0.0000	0.6355	0.2977