Journal of Organizational Behavior Review (JOB Review)

Cilt/Vol.: 6, Sayı/Is.: 1
Yıl/Year: 2024, Sayfa/Pages: 69-108
Başvuru Tarihi: 25.11.2023
Kabul Tarihi: 24.01.2024
Derleme Makale/Review Article

Structural Equation Modeling (SEM) for Social and Behavioral Sciences Studies: Steps Sequence and Explanation

Marwan GHALEB[1] Muhsin Murat YAŞLIOĞLU [2]

Abstract

Structural equation modeling (SEM) is one of the multivariate analyses that is used to test complicated research models, which include several constructs that have a group of independent and dependent variables with a series of hypothesized relations and associations between them. It starts with examining the collected data by conducting a data screening analysis and descriptive statistics. The scale used to measure the variables should be examined by conducting factor analysis (EFA & CFA) to make sure the data fits the research measurement model and to assess the scale's reliability, validity, and its level of fit to the collected data. The analysis of multivariate assumption should be reviewed then path analysis can be done for hypotheses testing and getting the final results. The final results have to be explained and interpreted based on the research's theoretical background and its literature foundation. This review paper explains these steps in detail for quantitative analysis students and other researchers who have basic knowledge of statistics, using simple words without diving deeply into statistics' details and its related formulas.

Keywords: *Structural equation modeling (SEM), Descriptive statistics, Factor analysis, Reliability & Validity, Path analysis.*

Sosyal ve Davranışsal Bilimler İçin Yapısal Eşitlik Modellemesi (YEM): Adımlar Dizisi ve Açıklaması

Öz

Yapısal Eşitlik Modellemesi (YEM), çeşitli yapıları içeren karmaşık araştırma modellerini test etmek için kullanılan birçok değişkenli analiz yöntemidir. Bu yöntemde, bir grup bağımsız ve bağımlı değişken arasındaki ilişkiler varsayılan ve ölçekleri belirli bir şekilde yapılandırılmıştır. Analiz, veri tarama analizi ve tanımlayıcı istatistikler yürüterek toplanan verilerin incelenmesiyle başlar. Verilerin ölçülmesi için kullanılan ölçek, faktör analizi (EFA & DFA) yöntemleri ile belirlenir ve verilerin araştırma ölçme modeline uygunluğu, ölçeğin güvenilirliği, geçerliliği ve toplanan verilerle uyum düzeyi değerlendirilir. Çok değişkenli varsayımların analizi gözden geçirilmelidir; daha sonrasında ise yol analizi yapılarak hipotezler test edilebilir ve nihai sonuçlara ulaşılabilir. Sonuçlar araştırmanın teorik altyapısı ve literatür temeline dayalı olarak açıklanlamalı ve yorumlanmalıdır. Bu makale, nicel analiz öğrencileri ve istatistik konusunda temel bilgiye sahip diğer araştırmacılar için, istatistiğin ayrıntılarına ve ilgili formüllerine girmeden basit kelimeler kullanarak bu adımları ayrıntılı olarak açıklamaktadır.

Anahtar Kelimeler: *Yapısal Eşitlik Modellemesi (YEM), Tanımlayıcı istatistikler, Faktör analizi, Güvenilirlik ve Geçerlilik, Yol analizi.*

[1] Ph.D. candidate, Istanbul University, Institute of Social Sciences, Department of Business Administration, Istanbul, Turkey, E-mail: marwan.m.ghaleb@gmail.com,
[2] Assoc. Prof., Istanbul University, Faculty of Business Administration, Department of Business Administration, Istanbul, Turkey, E-mail: muratyas@istanbul.edu.tr,

Introduction

Both social and behavioral science is a field of study that is continuously developing. Studying the relationships between different variables has become more complicated, as in real life it is very rare to find only one variable affecting another variable, actually, there is an unlimited number of variables affecting and affected by each other at the same time (Jöreskog & Sörbom, 1982). Each variable has a different role than the other, such as independent, dependent, mediating, and moderating variables. They can also be first-order, second-order, or control variables (Hair Jr et al., 2014). Studying different types of variables and their relations with each other leads to results that reflect reality in a better way.

Due to the development of social and behavioral sciences, sophisticated advanced theories have been developed to reflect the actual reality. Such theories need complex research models that include multiple observed variables and multiple relationships, which cannot be addressed with simple research models that have a limited number of variables, that’s why advanced statistical models are developed such as Structural equation modeling (SEM) (Jöreskog & Sörbom, 1982; Schumacker & Lomax, 2010).

As same as any other type of analysis in SEM the collected data should be examined and make sure that it fits the scale used to measure the variable, also the scale itself should be tested to make sure that it is appropriate for the research model under investigation, then the research hypotheses can be tested (Kline, 2011).

This paper targets quantitative analysis students and other researchers in social and behavioral science who have basic knowledge of statistics. It discusses what SEM analysis is in detail and shows how it can be conducted to analyze a research model related to social and behavioral studies using simple words and examples without diving deeply into statistics and its related formulas. It combines the needed information from books and other resources in one place hoping to save time and effort.

It aims to explain the steps of conducting an SEM analysis and their sequence for social and behavioral science studies and provides information about the purpose of each step and its related outputs. These steps can be categorized into three categories:

1. Preliminary analysis, which is done before starting any analysis. It includes data screening and descriptive statistics.

2. Factor analysis, which can be considered as the starting point of SEM. It includes Exploratory factor analysis (EFA), Confirmatory factor analysis (CFA), as well as reliability and validity assessment.
3. Path analysis (with its related assumptions), which includes the hypotheses testing.

1. What is Structural Equation Modeling (SEM)

SEM is an advanced statistical model that is capable of testing and estimating relationships among complicated research models (constructs) that include multiple variables (Hair Jr et al., 2014; Weston, 2006). It provides a quantitative analysis of the hypothesized research model created to address a phenomenon, this research model is presented by a set of variables that define a certain construct and their relations with each other (Schumacker & Lomax, 2010). It is used when a research model includes hypotheses with a series of independent and dependent variables in a specific order, that needs a group of multiple regression analyses to explain the relation and causal effects among them (Hair Jr et al., 2014; Jöreskog & Sörbom, 1982).

Researchers who are planning to use SEM must have foundation knowledge in statistical analysis such as correlation and regression (Schumacker & Lomax, 2010), they also have to define their research model with its related relationships in advance, then by using SEM they can test if the data they collected explain and reflect their research model or not (Weston, 2006).

Byrne (2010), defined SEM as "a statistical method to analyze a structural theory that explains some phenomenon, which represents a cause and effect (causal) process that creates observations on several variables related to the phenomenon, using a confirmatory and hypothesis testing approach". It includes two components: The first component is the measurement model, which is represented by the factor analysis, and the second component is the structural model which is represented by path analysis or in other words advanced multiple regression analysis (Hair Jr et al., 2014, Weston, 2006).

Byrne (2010) also, explained that there are two parts of the statistical analysis process represented by SEM: The first part represents the series of structural equations such as regression formulas, that are used to test the research cause and effect model, and the second part represents the research structural relations that can be drawn or pictured as a figured model that helps to present a quick understanding of the research model.

According to Schumacker & Lomax (2010), the goal of SEM is "to determine the extent to which the theoretical research model is supported by the sample data collected". It determines whether the data collected supports the research model in explaining the phenomenon based on a certain theory or not, otherwise, if there is no such support from the data collected, the research model can be modified or replaced and tested again.

Hair Jr et al. (2014), Byrne (2010) and Schumacker and Lomax (2010), explained that SEM helps in improving the understanding of complicated relationships in a research model by following scientific methods of testing the research model and its related hypothesis, it steps ahead of other multivariate analyses due to several points:

1. It uses a confirmatory approach for the data analysis rather than an exploratory one, and it helps in identifying the validity and reliability of the measurement tool (the scale and its related survey questions).
2. It helps in calculating the measurement error and takes it into consideration when presenting the final results.
3. It presents clearly, the latent variable (the variable which is measured) with its related observed variables (survey questions) and the role of each observed variable in measuring the latent one.
4. It helps in testing indirect effects such as a mediating effect, or moderating effect, it also helps in testing the research model when some variables are controlled.
5. Finally, it helps in group comparison when examining the same model and collecting the data from two different groups to compare.

On the other hand, according to Weston (2006), SEM is similar to other analyses such as correlation, regression, comparison T-test, and variance analysis ANOVA in some points:

1. They represent linear models.
2. Each one of them has its assumptions that should be met to conduct the analysis.
3. All of them help to understand hypothesized relationships based on a theory.

After understanding what is SEM analysis the next step is to know how it can be used for data analysis and hypotheses testing, starting with the preliminary analysis used to examine the collected data.

2. Preliminary Analysis

Before starting any hypothesis testing the collected raw data must be organized, examined, and evaluated, to make sure of its quality and that it does not include any possible errors, this is called the preliminary analysis (Jhangiani et al., 2019; Tabachnick & Fidell, 2013).

The preliminary analysis includes ensuring that the surveys are filed appropriately and there are no insufficient effort responses that may lead to outliers as well as a skewness or kurtosis distribution. It also contains descriptive statistics such as calculating the mean and standard deviation in addition to assessing the normal distribution (Jhangiani et al., 2019; Tabachnick & Fidell, 2013).

How to detect insufficient effort responses and data screening, in addition to descriptive statistics are discussed in detail in the following section:

2.1. Insufficient Effort Responses and Data Screening

When collecting data, especially in social and behavioral sciences, researchers mostly use surveys to collect a large amount of data with less effort, time, and cost (DeSimone et al., 2015). Recently the online methods of using surveys to collect data have made the data-collecting process much easier, accessible, and widespread, however, it also increased the potential of collecting low-quality data (Curran, 2016), as the researchers cannot observe the survey respondents directly (DeSimone et al., 2015).

The collected low-quality data is due to survey respondents who are not motivated to answer the survey accurately, they provide inconsistent, random, or careless responses (Huang et al., 2012). This low-quality data that comes from unmotivated respondents leads to errors when conducting the data analysis; such errors can be prevented if these responses were removed from the total data collected. (Curran, 2016; Huang et al., 2015).

Such low-quality responses are called insufficient effort responses, "which occurs due to a lack of motivation to comply with survey instructions and to correctly interpret item content" (Huang et al., 2015). Including such carelessly random responses affects the overall data set quality, which leads to unexpected problems during the analysis and questionable undesired results (Curran, 2016; Huang et al., 2015; Huang et al., 2012).

Insufficient effort responses must be detected because removing them helps to increase the quality of the data collected. That's why it is highly recommended to cond data screening before starting the analysis to find these responses, correct them if applicable (for example if some answers were left empty), or remove them from the overall data collected (Curran, 2016; Huang et al., 2012).

Huang et al. (2012), defined insufficient effort responses as "a response set in which the respondent answers a survey measure with low or little motivation to comply with survey instructions, correctly interpret item content, and provide accurate responses". Huang et al. (2015), added that they represent "a survey response set in which a person responds to items without sufficient regard to the content of the items and/or survey instructions".

There are several types of insufficient effort responses such as responses that indicate different answers to similar questions, or the opposite which is responses that indicate a similar answer to different questions in the survey (DeSimone et al., 2015). Some respondents provide the same answer to all the questions of a survey leading to an illogical response and, definitely, insufficient effort responses, also duplicating the collected survey responses to increase the sample number leads to insufficient effort responses (Huang et al., 2012).

Insufficient effort responses arise from respondents who do not have an interest in the survey or are not motivated to participate in the research (Meade & Craig, 2012). These responses affect the analysis outcomes such as mean and standard deviation (Curran, 2016), as well as they can negatively affect the measurement structure leading to the decrease of the scale (survey questions) reliability and its construct validity (Huang et al., 2015).

There are several methods to detect insufficient effort responses, some of them can be described as prevention methods that are implemented when designing the survey such as:

1. Detecting illogic responses using demographic questions, for example, a respondent who claims to have years of experience more than his age or a respondent who claims to have a high monthly salary while still a secondary student should be removed from the study (Huang et al., 2012).
2. Detecting the inconsistent answers to questions that measure the same concept such as answering strongly agree with the sentence "I am motivated at my workplace" and also strongly disagree with the statement "I am motivated in the place I am working in" (Huang et al., 2015).

The other methods can be used after collecting the data such as:

1. Detecting the respondents that have the same pattern of answers, as some of the respondents may repeat their first three or five responses rapidly to all the survey questions creating a certain pattern without reading them (Huang et al., 2015). This can be done by eye scanning; it may take time but it helps improve the data quality.
2. Using a statistical approach to find identical answers by calculating the standard deviation of each response to the survey question, a standard deviation with a value of zero or closer to zero indicates an identical answer to all the survey questions (Tabachnick & Fidell, 2013).
3. Creating the data distribution graphs of each variable helps to detect outliers and extreme responses (DeSimone et al., 2015; Meade & Craig, 2012; Tabachnick & Fidell, 2013).

There is one more method that is used when collecting the data using paper surveys, which is to observe the time spent in answering the survey questions, a respondent who completes the survey in less than the expected minimum time of answering may be considered an insufficient effort respondent (DeSimone et al., 2015; Huang et al., 2015).

Finally, it is also recommended to detect the responses that include questions left empty and remove them if such questions are a lot, or fill them using appropriate statistical methods if they are few (Tabachnick & Fidell, 2013).

2.2. Descriptive Statistics

Descriptive analysis helps in summarizing the collected data into numbers, these numbers have a meaning that directs the researcher to a conclusion (Fisher & Marshall, 2009). It starts by describing the main characteristics of the collected data which represent the targeted sample as a whole and its related subgroups (Thompson, 2009), such analysis is called descriptive statistics. It is defined as “the numerical and graphical techniques used to organize, present and analyze data” (Fisher & Marshall, 2009), they are represented by “numbers that summarize the data to describe what occurred in the sample” (Thompson, 2009).

It is one of the simplest analyses that can be performed and interpreted, it is the easiest method to summarize a data set, get a description of the targeted sample, and show its characteristics (Fisher & Marshall, 2009; Marshall & Jonker, 2010). Understanding the targeted sample characteristics helps to get an idea about the nature of the groups within the sample, compare the samples from one study to another, and provide the researchers with information that may influence their final conclusion (Thompson, 2009).

The main analyses used to describe a set of data are frequency distribution, mean, and standard deviation, in addition to some other analyses such as median, mode, and range (Thompson, 2009), the output of these analyses is a single number that has meaning when it is used to describe a data set (Livingston, 2004).

Frequency distribution represents the accumulated number of respondents (samples) that select each of the alternative responses to a survey question. It represents the number of cases for each category (response), and it provides an initial idea of the respondents' direction toward a survey question. It can be represented by numeric values, graphs, and percentages (Fisher & Marshall, 2009; Thompson, 2009).

Thompson, (2009), explained that frequency distribution is appropriate for describing quantity within categories of nominal or ordinal data such as demographic data. It helps in detecting data entry errors, that's why it is recommended to start with frequency distribution analysis when starting to analyze a data set.

The mean is one of the descriptive statistics outputs, it represents the calculated average of the frequency distribution (Fisher & Marshall, 2009; Marshall & Jonker, 2010). After getting the value of each response, the mean is calculated by dividing the total value of all the responses by the number of these responses (Fisher & Marshall, 2009; Livingston, 2004; Thompson, 2009). It is affected by the availability of any outliers in the data set, and it is more appropriate for the interval and ratio data (Thompson, 2009).

Another output of the descriptive statistics is standard deviation, which represents the spread of the data distribution (Fisher & Marshall, 2009; Marshall & Jonker, 2010), it shows the average distance from each of the responses (observations, scores, samples) to the mean (Fisher & Marshall, 2009; Thompson, 2009), in other words, it shows how closely the responses are distributed around the mean (Marshall & Jonker, 2010). It is considered as a measure of the variance of a data set (Marshall & Jonker, 2010; Thompson, 2009).

Standard deviation is appropriate to be calculated for continuous data (Fisher & Marshall, 2009). It also helps to detect outliers and identical answers, as a high value of a standard deviation indicates a big distance between the response and the mean, which in turn indicates the availability of outliers that can be found by further analysis (Thompson, 2009), on the other hand, a zero or close to zero value of a standard deviation for a response indicates identical responses to the survey questions, an average value of a standard deviation is more appropriate than a high or a low value.

Lee et al. (2015), explained that examining the mean and the standard deviation of a data set helps to know if the data is normally distributed or not by calculating the variation of the data and the distance of each observation (response, sample) from the mean.

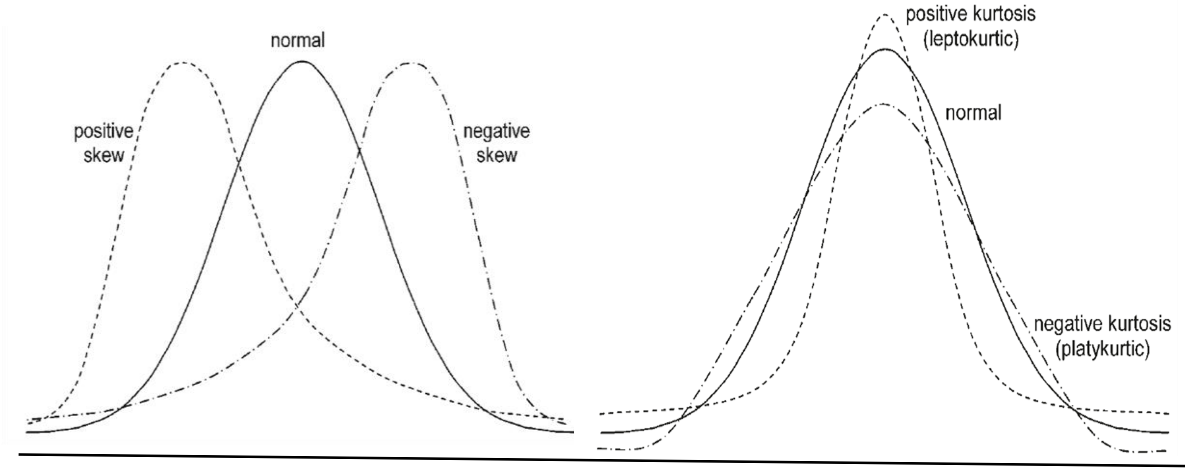
This leads to one method of descriptive statistics that is used to describe a collection of values (responses, observations), which is called the data distribution (Thompson, 2009). When describing the distribution of the values or the data around the mean, the values are either normally distributed or non-normally distributed.

The data is normally distributed when it takes a bell curve shape with two symmetric sides, and the middle of this bell shape is the center of the distribution (Thompson, 2009). The normal distribution is one of the main conditions for conducting parametric analysis such as correlation, regression, T-test, and ANOVA, however, due to outliers the perfect bell shape may not be achieved (Marshall & Jonker, 2010). The more data to be collected the better normality distribution curve is achieved.

A non-normally distributed data is represented by a skew or kurtosis curve shape. A skewed distribution is represented by a non-symmetric distribution around the mean, a positive skew occurs when the distribution curve is moved to the left, which means that the observations' values are less than the mean, on the other hand, a negative skew occurs when the distribution curve is moved to the right, which means that the observations' values are more than the mean (Kline, 2011; Tabachnick & Fidell, 2013). A kurtosis distribution is represented by a distribution with a higher peak that indicates a positive kurtosis, or a lower peak that represents a negative kurtosis (Kline, 2011; Tabachnick & Fidell, 2013). Parametric analysis is not appropriate for non-normally distributed data as it is analyzed by non-parametric analysis tests (Marshall & Jonker, 2010). Figure 1 compares the normal and non-normal distribution curves.

Figure 1

Normal and Non-normal Distribution Curves.



Source: (Kline, 2011).

Before starting a factor analysis, it is important to make sure that the distribution of each question response is not skewed or kurtosised, as such distribution indicates a low factor loading of a question. Skewness and kurtosis can be calculated numerically by statistics software such as SPSS. Usually, a skewness and kurtosis value between 1 and -1 are considered in the most appropriate range (Hair et al., 2022; Schumacker & Lomax, 2010; Tabachnick & Fidell, 2013), values that are more than 1 and -1 are considered moderate skewness and kurtosis that can also be accepted for factor analysis, but a skewness value more than 3 and -3 or a kurtosis value more than 8 and -8 are considered extreme and they are not accepted for factor analysis (Griffin & Steinbrecher, 2013; Kline, 2011).

To conclude, descriptive statistics helps to describe the data set, find potential trends in the demographic data of the survey respondents, or help in interpreting the research's final results. However, it cannot be used for causal analysis, to explain relations between variables, or to generalize the results from the sample tested to the entire population, as this needs the use of inferential statistics such as hypotheses testing (Marshall & Jonker, 2010).

Before the hypotheses are tested and after the preliminary analyses are done, the scale must be evaluated, this is done by factor analysis (Fisher & Marshall, 2009; Marshall & Jonker, 2010).

3. Factor Analysis

To understand the nature of factor analysis it is essential to know what is reflective and formative measurement models. In a reflective measurement model, the latent variable affects the observed variables, any changes in the latent variable lead to changes in its observed variables, and losing any observed variable will not affect the latent variable, however, the latent variable cannot reflect or present itself it must be reflected and presented by its related observed variables. On the other hand, the formative measurement model is completely the opposite of the reflective measurement model (Brown, 2015; Coltman et al., 2008; Hair Jr et al., 2014).

Factor analysis is used to examine the relationship between a group of observed variables and latent variables in a reflective measurement model (Byrne, 2010). The group of observed variables' job is to reflect and present the latent variable, which cannot be presented by itself or by only one observed variable (Hair Jr et al., 2014). They are represented by observations, scales, survey questions, or any other measuring items (Beavers et al., 2013). To understand the latent variable, the covariance between the group of observed variables is tested to generate a construct (factor) that includes the observed variables that are highly correlated with each other, this factor is used to represent the latent variable (Byrne, 2010; Henson & Roberts, 2006). So based on their covariance, the group of observed variables is reduced from a larger set of variables to a smaller one that appropriately represents and explains the latent variable, with the least amount of information loss (Hair Jr et al., 2014; Henson & Roberts, 2006; Yong & Pearce, 2013).

Factor analysis helps to assess the integrity of the scale used to measure the variables (latent variables), it also helps in discovering the underlying structure (construct or factor) in the collected data, and to what extent this structure represents the variables under measurement (Henson & Roberts, 2006). Researchers use factor analysis when constructing a questionnaire to measure a variable, as it leads to reducing the questions to a manageable amount that represents the variable as much as it can, which can help in collecting the data and reduce the unengaged responses (Field, 2013).

There are two types of factor analysis used when developing a scale, they are Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) (Byrne, 2010; Orcan, 2018). EFA is used first, when the relationship between the observed variables is unknown, as it helps to explore their underlying structure based on the collected data, on the other hand, CFA is used when the relationship between the observed variables is known, as it confirms the availability of an already known structure (explored in EFA) in a newly collected data, and shows whether such data is appropriate for path analysis and hypotheses testing or not (Orcan, 2018; Yong & Pearce, 2013). EFA and CFA in addition to assessing the scale’s internal consistency, reliability, and structural validity are discussed in detail following sections:

3.1. Exploratory Factor Analysis (EFA)

EFA is a statistical technique that is used for finding the underlying structure (factor) among observed variables and its relation with the latent variable when there is no knowledge about the links between the observed variables (questions or items) and the latent variable, it shows that latent variable is there, and which questions or item is related to which variable (Byrne, 2010; Hair Jr et al., 2014; Henson & Roberts, 2006; Orcan, 2018).

It is commonly used when developing a new scale as it helps to discover the existing structure of the observed variables based on the collected data and which structure is related to which latent variable (Orcan, 2018). Doing an EFA means that a researcher is exploring how and to what level the observed variables are connected to their underlying structures (factors), as there is no previous knowledge of whether they are appropriate to measure the latent variable or not (Byrne, 2010; Orcan, 2018).

In EFA the relationship between the observed and the latent variables is represented by the factor loadings, as the higher the loading the stronger the relationship between them (Byrne, 2010). When it is used to search for structure among variables, the factor loading helps in finding that structure, as the observed variables (items or survey questions) with low loading can be excluded, based on that a smaller set of observed variables that are more interrelated with each other will remain (Hair Jr et al., 2014; Yong & Pearce, 2013).

Reducing the observed variables based on factor loading helps in simplifying the structure in a way that makes it easier to understand and interpret the relations with a minimal loss of information (Hair Jr et al., 2014; Yong & Pearce, 2013). The concept of EFA is basically, that there is a number of common latent structures (factors) to be explored in a dataset, and the goal is to find the minimum number of observed variables that have covariance with each other to represent a latent structure based on their factor loading (Yong & Pearce, 2013).

EFA is also, used when adopting an original scale to a new language as due to translation there is a risk of losing the original meaning of the scale items (survey questions), so doing an EFA helps in finding the structure of the variables based on a new dataset after the translation (Orcan, 2018), it is also used when combining different scales together to create a new combined scale to test a combined set of variables in a research model.

To conduct an EFA analysis the observations (or respondents to survey questions) should be more than 50, the amount of observations for an EFA can also, be determined by five times the number of variables (number of variables multiplied by 5) (Hair Jr et al., 2014). However, the commonly accepted amount of observation for an EFA is 100 observations (Beavers et al., 2013). There are several methods of conducting factor analysis, the most appropriate method for EFA is the method of principal axis for factor extracting (Field, 2013; Hair Jr et al., 2014).

As same as any other multivariate analysis, factor analysis relies on some assumptions, such as multivariate normality, linearity between factors and their related observed variables, homoscedasticity, as well as not having outliers and extreme multicollinearity (Field, 2013; Hair Jr et al., 2014; Yong & Pearce, 2013; Beavers et al., 2013). However, considering multivariate normality as an assumption of factor analysis is debatable because in factor analysis the relation between variables is tested without defining which variable influences the other, so multivariate normality is not assumed for some extraction methods of factor analysis such as principles components and principles axis, on the other hand, maximum likelihood assumed multivariate normality (Beavers et al., 2013; Brown, 2015). These assumptions will be discussed in detail later in this paper.

The first step in doing an EFA is to make sure that the collected data is appropriate for factor analysis, this is done by calculating the percentage of Kaiser-Meyer-Olkin (KMO), which shows if the observations are enough for factor analysis, and whether they have shared variance or not, the accepted percentage of KMO is 0.70 or more (Beavers et al., 2013). Bartlett's test of sphericity also should be done at the start of an EFA to make sure that structures (factors) can be extracted from the collected observations (data), and that there is a correlation between them, it has to be less than 0.05 (Beavers et al., 2013; Hair Jr et al., 2014). Getting the accepted results of these two tests is the key to opening the door to starting an EFA.

The next step is to calculate the communalities, which can be done by SPSS software. Communalities represent the level of shared variance between an observed variable and all unobserved variables included in the analysis. It is used to indicate the level of explanation presented by each observed variable toward all latent variables (factors), an observed variable with a shared variance value less than 0.50, doesn't represent enough explanation from the latent variables, and it is subjected to elimination from the analysis in the next step of evaluating the factor matrix (Hair Jr et al., 2014; Henson & Roberts, 2006; Yong & Pearce, 2013).

After that, the factor matrix (if only one factor is extracted), or the rotated factor matrix (if more than one factor is extracted) is used to evaluate the factor loading, which is used to measure the level of contribution from the observed variables to the factor (latent variable) (Yong & Pearce, 2013). Based on factor loading a structure can be found when a group of observed variables has a high factor loading on a factor (or a single latent variable) (Field, 2013; Hair Jr et al., 2014). The cut-point percentage of an accepted factor loading differs based on the number of observations (samples or participants), the fewer the number of observations, the higher the factor loading cut-point percentage of acceptance. Table 01, which is taken from Hair Jr et al. (2014), shows the factor loading acceptance cut point based on sample size (observations size) at the significant level of 0.5.

Table 1

EFA factor loading acceptance cut point based on the sample size at the significant level of 0.5.

Sample / Observations size	**Factor loading acceptance cut point**
350	0.30
250	0.35
200	0.40
150	0.45
120	0.50
100	0.55
85	0.60
70	0.65
60	0.70
50	0.75

Source: (Hair Jr et al., 2014)

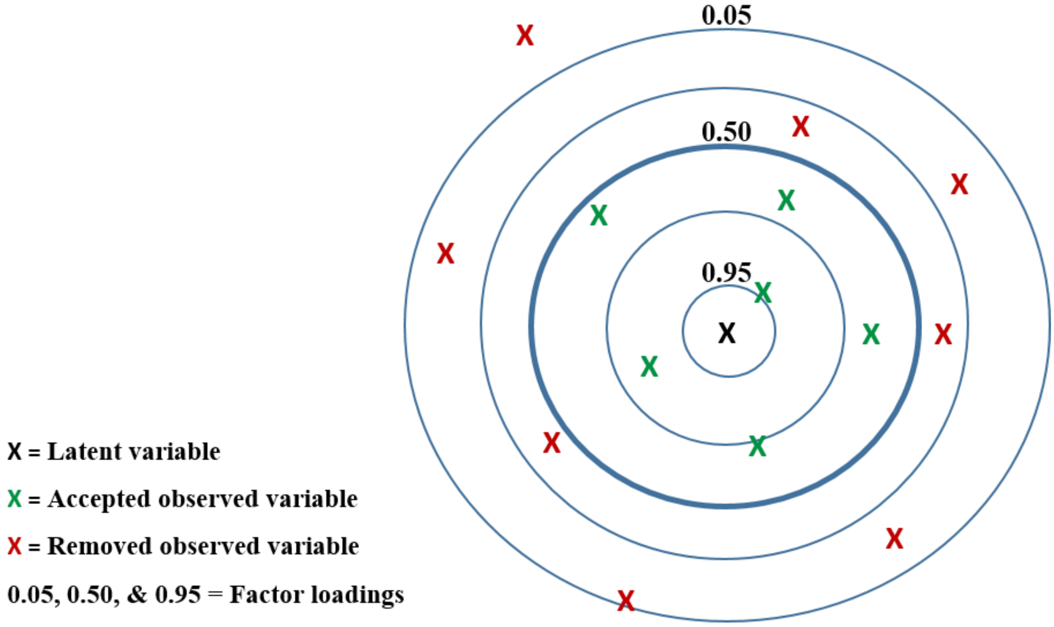
However, based on the number of observations the commonly used factor loading percentage is 0.50 or more (Hair Jr et al., 2014). Observed variables that have a factor loading less than the accepted cut point or have multiple factor loading on more than one latent variable should be removed one by one because removing one observed variable leads to changes in the factor loading of the others, some factor loading may improve and the other may not (Hair Jr et al., 2014). Finally, an accepted structure that represents a latent variable should have at least 3 observed variables (Yong & Pearce, 2013).

To simplify the understanding of the EFA, consider the overall survey questions are divided into groups, each survey question within a group has a task of measuring a certain variable. The questions that succeeded in their task and got a high factor loading should stay in the group and represent the variable, but the other questions that didn't should be removed, additionally, the questions that were jumping and dividing their factor loading between groups should also, be removed.

Figure 2 visually summarizes the concept of accepting or removing an item based on factor loading at the acceptance cut-point of 0.50. It simulates EFA with the game of hitting the target with arrows, in EFA the survey questions are like the arrows and the latent variable that needs to be measured is the target. Within the cut-point, the more the question hits closer to the target the more it is likely to be accepted to measure the variable.

Figure 2

Factor Loading Explanation



Source: Developed by the author.

Finding the underlying structure after conducting an EFA means that the researcher becomes aware of the relationship between the observed variables, so the next step is to confirm the found structure and make sure that all the latent variables with their observed variables, which are related to the explored structure, are fitting together, and they are appropriate to have hypothesized paths between them, this is done by conducting a CFA (Hair Jr et al., 2014). However, it is suggested that after EFA analysis and removing the low-factor loading questions, the Cronbach Alpha value for each latent variable can be calculated, to check the internal consistency of the scale.

3.2. Cronbach Alpha (α) for Internal Consistency (Reliability) Assessment

Reliability represents the internal consistency of a scale's measurement values when doing the test again and again under the same conditions (Ercan et al., 2007; Tavakol & Dennick, 2011). In social and behavioral sciences, scales (survey questions) are used several times on different groups of participants to measure a certain variable, which allows comparison between groups, so the consistency of the scale should be evaluated to make sure that everyone understood the questions and the results can be used for further comparison or any other statistical testing (Tavakol & Dennick, 2011). A reliability evaluation, which represents a scale's consistency evaluation, has a value between 0 and 1 (Ercan et al., 2007), and can be calculated using Cronbach's alpha analysis of reliability assessment (Vaske et al., 2017).

Cronbach's alpha is a useful commonly used analysis to examine the reliability of a scale (Bonett & Wright, 2015; Brown, 2002), it evaluates the consistency of the items (survey questions) of a scale (Vaske et al., 2017), it is widely used in social and behavioral sciences studies, especially when using a Likert scale in collecting the data (Bonett & Wright, 2015; Ercan et al., 2007). It works by estimating the variance available in a collected data set to see whether it is consistent (has a consistent systematic pattern) or not, such estimation has a value between 0 indicating no consistent variance, and 1 indicating that all the variance is consistent (Brown, 2002). The variance represents to what level the data differs from its mean or the distance between the data and its related mean.

Usually, a Cronbach's alpha value between 0.65 and 0.80 is considered acceptable to say that a scale is reliable (Vaske et al., 2017), in social and behavioral sciences a value of 0.70 and more is an indicator of good reliability, however, in some cases, 0.60 is considered as an accepted indicator of reliability (Hair Jr et al., 2014; Vaske et al., 2017). This value is affected by the number of scale items (survey questions) as well as the amount of data collected, the more items and data the higher the value (Brown, 2002; Tavakol & Dennick, 2011; Vaske et al., 2017). However, if the test has more than one variable (concept) Cronbach's alpha should be calculated for each variable separately to find the reliability of the items related to that variable, as it is incorrect to calculate an accumulated Cronbach's alpha one time for all the scale's items regardless of their related variable (Tavakol &

Dennick, 2011). It is recommended that Cronbach's alpha should not be too high as this indicates that some items are highly correlated to the level that they can be considered as one item. In this case, it is suggested to reduce the number of items (factor analysis can be used for that), some authors say that it should not exceed 0.90 (Tavakol & Dennick, 2011).

One of Cronbach's alpha assumptions is that the items are positively correlated, however, in some cases, a negative Cronbach's alpha value may appear, which indicates a negative correlation between the items. This indicates that an item or more needs to be re-coded, especially if it is a negative item (a negative survey question that needs an upside-down coding). It also indicates that there might be inconsistent unengaged responses from the respondents of the survey questions (Vaske et al., 2017). Additionally, a low value of Cronbach's alpha might be due to a low interrelatedness between the items, the sample is not homogeneous, and the survey questions are not enough to measure their related variable (Tavakol & Dennick, 2011).

Cronbach's alpha doesn't indicate consistency in different time periods, so a researcher should not rely on a published reliability value of a certain scale, as it has to be calculated again each time the scale is used, this is called the test-retest reliability method (Brown, 2002; Tavakol & Dennick, 2011; Vaske et al., 2017).

Finally, based on the correlation assumption of Cronbach's alpha, it can be used to calculate the measurement error by squaring its value and then subtracting the result from 1, for example, if the Cronbach's alpha value is 0.75 so the error will be 0.44 (0.75*0.75=0.56 then 1-0.65=0.44) (Tavakol & Dennick, 2011). The next step is to confirm the factors explored in the EFA by conducting a CFA.

3.3. Confirmatory Factor Analysis (CFA)

In EFA the collected data is the primary source of the factor structure, based on statistical analysis the underlying pattern of the data is explored and used to define the latent variables, their related observed variables, and the percentage of factor loading of each observed variable, which represents the explored factor structure (Hair Jr et al., 2014). On the other hand, CFA is used when the factor structure is already known and the research model is defined, as well as there is an idea about each latent variable and its related observed variables (Byrne, 2010; Orcan, 2018; Suhr, 2006).

In CFA, the relation between the latent variables and their observed variables is assumed based on a certain theory that has a proven factor structure. This structure confirms the appropriateness of the collected data to the defined research model (Byrne, 2010; Orcan, 2018; Suhr, 2006). This leads to finding out that the key difference between EFA and CFA is that CFA needs a defined conceptual framework (theory) to direct the evaluation of the factor structure for the collected data (Brown & Moore, 2012; Orcan, 2018; Schumacker & Lomax, 2010).

CFA is defined as a statistical technique that is used to examine a theoretical factor structure related to a set of observed variables, to find out if the actual collected data confirms the defined theoretical model or not (Hair Jr et al., 2014; Schumacker & Lomax, 2010; Suhr, 2006).

When creating a new scale and after exploring the factor structure based on the data collected during the EFA stage, CFA is used to confirm the explored factor structure (The factor structure is identified after EFA) (Brown & Moore, 2012), this helps to assess the structured validity of the explored factor structure and confirms the relationship between the observed variables and their related latent variable in the scale under development (Orcan, 2018), this should be done using a different data set other than the one used in the exploring stage (Schumacker & Lomax, 2010).

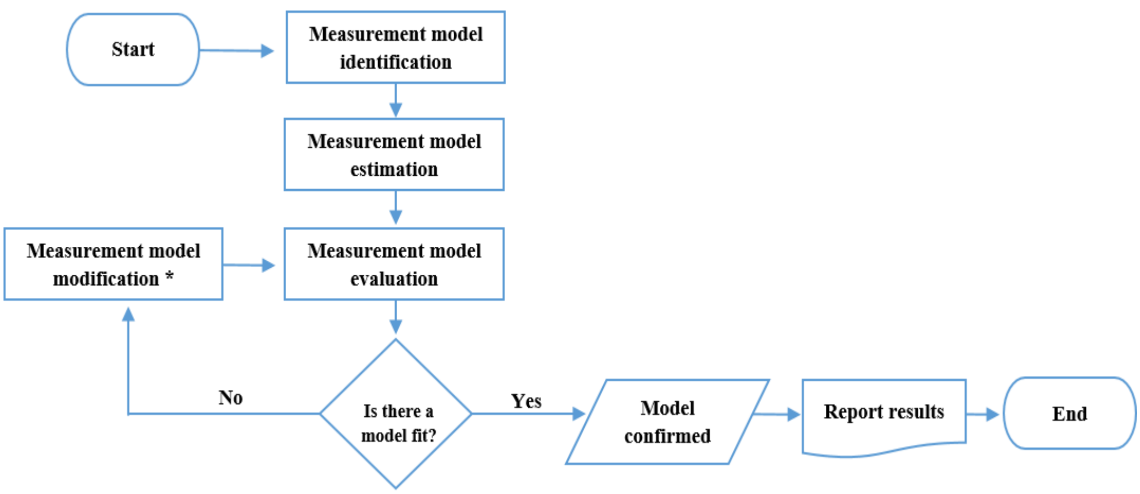
In the SEM analysis the CFA step is called the measurement model stage (Byrne, 2010) because, in this stage based on prior knowledge of a theory or an empirical study, a researcher has to define the number of variables (latent variable) in the research model and propose survey questions (observed variables) that measure each one of them in advance (Brown & Moore, 2012; Hair Jr et al., 2014; Suhr, 2006). In this stage, no statistical methods are used to identify the structure but they are used to confirm whether the collected data fit the proposed measurement structure or not (Hair Jr et al., 2014; Schumacker & Lomax, 2010). It helps to evaluate the compatibility of the proposed measurement model with different groups of data (Brown & Moore, 2012).

As explained, CFA is led by a theory, this theory is called the measurement theory. According to Hair Jr et al. (2014), this theory proposes how to measure a latent construct that is not measured directly; it hypothesizes a relation between the latent variable and a group of observed variables. The observed variables' job is to represent and measure the latent variable within the research model. The measurement theory is combined with the structure theory, which proposes the relation between the various variables in the research model, to identify the final SEM research model and test the research hypothesis. To prove a measurement theory, it should be tested using CFA several times on different data sets and compare the consistency of the results, this provides evidence of the measurement model's stability and its applicability to be generalized.

CFA steps start with model identification then model estimation, and finally, model evaluation (model fit evaluation), based on the evaluation results a model modification step might be done. These steps are presented in Figure 3. followed by a detailed explanation of each step.

Figure 3

CFA steps.



* When CFA is modified it turns into an Exploratory SEM analysis which is used for further analysis and hypothesis testing.

Source: Developed by the author, based on (Hair Jr et al., 2014; Miyao & Maruyama, 2012).

3.3.1. Measurement Model Identification

The first step when conducting a CFA is to identify the measurement model, this is done based on a measurement theory, this theory represents the foundation of identifying the number of factors (latent variables), the measurement variables (observed variables, survey questions), and which variable loads on which factor (Brown & Moore, 2012; Hair Jr et al., 2014). This step shows how the latent variable is going to be measured, and which measurement scale is going to be used, however unlike EFA in CFA prior information is needed to define the construct and set estimations tools for each factor in advance (Brown & Moore, 2012; Hair Jr et al., 2014).

There are some rules to follow when identifying a model in CFA, Hair Jr et al. (2014) and Byrne (2010) explained that: Observed variables are free to load on a factor based on the measurement theory, however, they are not allowed to have zero loading or to load on more than one factor. This means that an observed variable must be assigned to one latent variable as one survey question cannot be used to measure more than one variable. Finally, each factor (latent variable) should be identified by at least three measurement variables (observed variables).

Diagrams are used to visually represent a CFA measurement model (especially when using analysis software such as AMOS), rectangles represent the observed variables, and ovals or ellipses (flattened shape circles) represent the latent variables. These two shapes are connected by an arrow that represents the factor loading from the latent variable to the observed variable. Each observed variable should be linked with an error term represented by a circle that includes the letter “e”. Finally, the relationship between the latent variables should be represented by double-sided arrows (Hair Jr et al., 2014; Kline, 2011). When running and evaluating the CFA model each shape will have a number that is used to identify model fit based on the collected data (Byrne, 2010).

3.3.2. Measurement Model Estimation

The next step after identifying the measurement model is to estimate the relationship between the latent variables (factors, constructs) and their observed variables (measurement variables, survey questions), this is called the model estimation step. In this step, the measurement theory that was used to identify the model is compared with the actual measurement of the model which is done based on the collected data, to find out if the theory fits the collected data or not (Hair Jr et al., 2014). The estimation method used in CFA is the maximum likelihood, which helps to find the best structure (measurement model) that fits the collected data, by maximizing the probability likelihood function, so that the observed variables are most probable to measure the latent factor (latent variable) under the identified measurement model (Yaşlıoğlu & Toplu Yaşlıoğlu, 2020).

Factor loading between the latent variable (construct) and its related observed variable (items) is used to estimate their relationship. High loading is expected to be found, it is represented by the standardized loadings estimates and should be between 0 and 1. Factor loading more than 0.5 to less than 0.7 can be taken into consideration if it does not affect reliability and validity negatively; an item with a loading less than 0.5 should be removed from the model. A factor loading around 0.7 or more is considered an ideal loading that indicates a strong relationship, however, a factor loading of 1 or more indicates some problems with the collected data (Hair Jr et al., 2014; Hamid et al., 2017).

In the CFA model estimation step, the R-squared (R^2) for each observed (measured) variable is also estimated. It also represents a value between 0 and 1, the higher the value the better the observed variable variance is explained by the latent variable, which means the better the relation between the observed variable and its related latent variable (Hair Jr et al., 2014). These values help in producing the predicted variance-covariance matrix of the measurement model which is used to estimate the residuals (Brown & Moore, 2012).

Generally, residuals represent the variation between each observed value and the observed mean of the collected data, in other words, they represent the differences between the observed actual covariance and the estimated covariance in the variance-covariance matrix. High residual values indicate a model fit issue, as the smaller the residual values the better the model fit (Hair Jr et al., 2014; Schumacker & Lomax, 2010). A residual value of 2.5 or less does not indicate an issue in the model fit, but a value of 4 or more indicates fitting issues in the model. Such model fit issues are represented by an unaccepted level of error which may lead to removing the item that causes this error. Finally, a residual value between 2.5 and 4 is not preferable but may not lead to any changes in the measurement model (Hair Jr et al., 2014).

Values and outputs of this step give an overall view of the measurement model and the relationship between the observed variables and their latent variable (factor) and lead to the next step of evaluating to what level the collected data fit the measurement model.

3.3.3. Measurement Model Evaluation (Is There a Model Fit?)

In the model evaluation step the model goodness of fit is evaluated, to find out whether the collected data fit the measurement model or not (Byrne, 2010). Getting a fit model in CFA means that the collected data is consistent with what is going to be measured, the appropriate measurement theory is selected to measure the research model’s variables, and it can be used in other similar researches with different data. On the other hand, getting a fit model doesn’t mean that the model is valid, but it is an indicator to move to further procedures to calculate the structured validity (Yaşlıoğlu & Toplu Yaşlıoğlu, 2020). The number of variables (factors) in the measurement model has an impact on the model fit values, the simpler the model and the fewer variables it has the higher the possibility of getting better model fit values, also normality of the collected data indicates a possibility of getting a better model fit (Yaşlıoğlu & Toplu Yaşlıoğlu, 2020).

The first indicator of a fit model is chi-square with its degree of freedom and its related P-value (Byrne, 2010). A chi-square value that is close to zero with a related P-value of more than 0.05 indicates that there is no difference between the expected variance-covariance matrix (of the measurement model) and the observed (actual) variance-covariance matrix achieved based on the collected data, which indicates a good fit (Alavi et al., 2020; Schumacker & Lomax, 2010; Suhr, 2006). This may be achieved with the minimum number of variables and the less amount of data, yet is not logical when conducting social or behavioral science research, that's why there are several alternative statistics values that indicate the model fit (Hu & Bentler, 1999; Suhr, 2006), such as:

- **CMIN/DF (Chi-Square Mean / Degree of Freedom):** Due to the sensitivity of chi-square to the sample size and the number of variables, dividing it by the degree of freedom is proposed to reduce such sensitivity. A value of CMIN/DF between 0 and 3 indicates a good model fit (Hooper et al., 2008; Hu & Bentler, 1999).
- **GFI (Goodness of fit index) and RMR (Root-mean-square residual index):** GFI and RMR are values between 0 representing no fit and 1 representing a perfect fit. A value of GFI around 0.90 to 0.95 represents a good fit, values between 0.80 and 0.90 could be considered an acceptable fit based on the number of variables and the amount of collected data (Schumacker & Lomax, 2010). A value of RMR less than 0.08 indicates a fit model and a value less than 0.05 indicates a fit model for a small amount of data. Generally, the accepted value of RMR is between 0.05 and 0.08 based on the amount of data collected. Using GFI and RMR to assess the model fit is not preferable (Yaşlıoğlu & Toplu Yaşlıoğlu, 2020).
- **Baseline comparisons fit indicators:** Since the model fit idea is based on comparing the variance-covariance matrix of the measurement model and the actual variance-covariance matrix of the collected data, three comparison-based indicators are used to assess the model fit. They are NFI (Bentler–Bonett normed fit index), TLI (Tucker–Lewis index), and CFI (Comparative fit index). All of them are values between 0 representing no fit and 1 representing a perfect fit. Based on the number of variables a value of 0.90 or above for each one of them indicates a fit model. CFI

is the more commonly used compared to the other two indicators (Hu & Bentler, 1999; Hooper et al., 2008; Schumacker & Lomax, 2010; Suhr, 2006; Yaşlıoğlu & Toplu Yaşlıoğlu, 2020).

- **RMSEA (Root Mean Square Error of Approximation) and PCLOSE (Root Mean Square Error of Approximation associated p-value):** Unlike the other indicators, RMSEA is a value between 0 representing a perfect fit, and 1 representing no fit, it shows how well the measurement model fits the variance-covariance matrix of the collected data. A value of RMSEA less than 0.08 represents a model fit, however, when having a larger sample size, it should be less than 0.05 (Schumacker & Lomax, 2010; Yaşlıoğlu & Toplu Yaşlıoğlu, 2020;). Generally, a value between 0.08 and 0.06 or less is generally accepted (Hooper et al., 2008; Hu & Bentler, 1999; Suhr, 2006). On the other hand, PCLOSE is a probability value that is linked with the RMSEA; it tests the hypothesis that RMSEA is equal to 0.05 (Kenny, 2020). Since a good fit needs a value of RMSEA less than 0.05, so the target of PCLOSE to indicate a good fit is to be more than 0.05 and reject that hypothesis (Hu & Bentler, 1999). There is a negative correlation between RMSEA and PCLOSE the more RMSEA value is the less the PCLOSE value will be. PCLOSE is more representative when there is a large sample and the RMSEA cut point is 0.05.

There is no rule to follow when selecting which model fit indicator to report, however, sample size and the number of variables help the researcher to determine which one to report, CMIN/df, CFI, RMSEA, and PCLOSE are mostly used (Yaşlıoğlu & Toplu Yaşlıoğlu, 2020).

3.3.4. Measurement Model Modification

After evaluating the model, if the data collected does not fit the measurement model and there is no model fit, the model can be modified and reevaluated (Schumacker & Lomax, 2010). Modification is needed when there is low goodness of fit (low model fit), high standardized residuals, and low factor loading (Brown & Moore, 2012).

Modification indices are part of the CFA outputs, they are calculated for possible unestimated relationships in the measurement model, and show whether the level of Chi-square decreases, if a covariance is added to the error terms of two observed variables related to one latent variable. No covariance can be made between two observed variables' error terms if each one of them is related to a different latent variable (Hair Jr et al., 2014; Yaşlıoğlu & Toplu Yaşlıoğlu, 2020).

According to Brown (2015) and Hair Jr et al. (2014), modification indices are sensitive to the number of observations (sample size), however usually in practice modification indices with a rounded value of 4 or higher are recommended to be modified. Modifications are recommended to be done one by one, starting with the higher value, and to be stopped once the model is fit, even if there are some more modification indices suggested. The next step if the model is not fit after doing the covariance between the error terms based on the modification indices, is to remove the questions (observed variables) with low factor loading one by one until a model fit is achieved, however, if more than 20% of the scale questions are deleted it is recommended to collect the data again and evaluate the measurement model with a different data set.

The target of modifications is to improve the measurement model not to change it, that is why it is recommended to have the minimum amount of modification (Hair Jr et al., 2014). Modifying a measurement model gives it an exploratory nature due to the minor changes done to fit the data, it is still a confirmatory factor analysis but with an exploratory nature, which can be relied on to calculate the reliability, validity, and later on to test the hypothesis. After modification, CFA becomes exploratory SEM analysis which is used for further analysis and hypothesis testing (Byrne, 2010; Kline, 2011; Yaşlıoğlu & Toplu Yaşlıoğlu, 2020).

After conducting the CFA, the researcher can assess the model's validity. Unlike the EFA which is used to identify the factors, CFA is used to confirm the factors and the validity of the measurement model (Hair Jr et al., 2014). The next step is to evaluate the composite reliability and the construct validation represented by the convergent and discriminant validity (Brown & Moore, 2012).

3.4. Assessing Reliability and Validity

In addition to the role of CFA in making sure that the collected data fit the measurement model, it is also used to examine the efficiency of the tool (survey questions, measurement model) used to measure a certain variable (latent variable, factor), this is done by to assessing the model's reliability and validity (Schumacker & Lomax, 2010). Once it is found that the measurement model is fit, reliable, and valid, then it can be used in further analysis and hypotheses testing (Byrne, 2010). Results are meaningless if reliability and validity are not assessed (Byrne, 2010; Schumacker & Lomax, 2010). CFA is used to assess the composite reliability and the construct validity.

Reliability, as explained earlier is defined as the degree to which the measurement is free from error (Muijs, 2004; Thanasegaran, 2009), it indicates the consistency of the results if the same measurement is used with a different group under the same conditions (Fitzner, 2007; Hair Jr et al., 2014; Sürücü & Maslakçı, 2020;). When evaluating which measurement tool to select for measuring a certain variable it is always recommended to use the one that proved high reliability (Hair Jr et al., 2014).

Cronbach alpha is most commonly used to assess reliability, however in SEM analysis composite reliability (CR), which is also called construct reliability, is used (Hair Jr et al., 2014). It represents the internal consistency between the latent variable (factor, construct) and its related observed variables in the measurement model (Schumacker & Lomax, 2010; Sürücü & Maslakçı, 2020). To conclude that a measurement is reliable, CR should be 0.70 or above (Hamid et al., 2017; Muijs, 2004). Assessing reliability is not enough as a reliable measurement tool may not be valid, that is why validity is important to be assessed (Sürücü & Maslakçı, 2020; Thanasegaran, 2009).

Validity is defined as the level at which a measurement model (tool, instrument) correctly represents and accurately measures the variable that needs to be measured (Fitzner, 2007; Hair Jr et al., 2014; Heale & Twycross, 2015; Schumacker & Lomax, 2010). It leads to answering a simple question, is the researcher measuring what is willing to be measured by the selected measurement model, and to what level the measurement model is accurately doing its job (Muijs, 2004; Sürücü & Maslakçı, 2020; Thanasegaran, 2009). It is all about how the latent variable is represented by the appropriate observed variables (survey questions) and to what level these observed variables define the latent variables and are used to shape an opinion about their related latent variable (Hair Jr et al., 2014; Schumacker & Lomax, 2010). It should be taken into consideration in the early stages of planning a research, by fully understanding the research topic, defining the variables appropriately, shaping the research model and hypothesis correctly, targeting the convenient population, selecting the sample using an appropriate method, and finally collecting the appropriate data to be measured by the appropriately selected measurement model (Hair Jr et al., 2014; Sürücü & Maslakçı, 2020; Thanasegaran, 2009). There are several types of validity, in this paper, construct validity is discussed, as it represents one of the outputs of the CFA.

Construct validity shows to what extent the measurement tool agrees with the measurement theory and whether the latent variable is actually measured based on the theoretical relationship with its related observed variables (Fitzner, 2007; Thanasegaran, 2009). It represents the level of accuracy in which a construct structure (latent factor) is measured by the selected measurement items (observed variables) based on the overall theoretical measurement model, which means that all the observed variables should load on their related latent factor (Byrne, 2010; Hair Jr et al., 2014). It includes convergent validity and discriminant validity (Sürücü & Maslakçı, 2020):

- **Convergent validity** represents the level of relationship between the observed variables themselves and their relationship with their latent factor (Sürücü & Maslakçı, 2020). To achieve convergent validity, the indicators (measures, observed variables) used to represent a latent factor should be interrelated and correlated with each other in the first place, then they have to be correlated with their latent factor only as there must be no or at least little correlation between those measures and other unrelated latent factors (latent variable) (Brown & Moore, 2012; Fitzner, 2007; Hair Jr et al., 2014; Schumacker & Lomax, 2010; Thanasegaran, 2009). This means that these indicators should have a high level of variance between them so that they can correlate with each other otherwise they will be considered as one same item (Hair Jr et al., 2014). High factor loading from the observed variables to their latent factor indicates the availability of convergent validity (Hair Jr et al., 2014). It is determined by calculating the average variance extracted (AVE) which is a value between 1 and 0 and should be equal to or more than 0.50 to conclude that convergent validity is achieved. It is also recommended that AVE should be less than CR (Gefen & Straub, 2005; Hair Jr et al., 2014; Sürücü & Maslakçı, 2020).
- **Discriminant validity** shows to what level different latent factors in a research model are separated from each other, this is represented by having no or the minimum correlation between them. This also means that the observed variables of each latent factor should be separated from other factors that are not related to them (Brown & Moore, 2012; Byrne, 2010; Hair Jr et al., 2014; Sürücü & Maslakçı, 2020). It is determined by:

    1. Calculating the maximum shared variance (MSV) which is a value between 1 and 0 and should be less than AVE to conclude that the discriminant validity is

achieved (Gefen & Straub, 2005; Schumacker & Lomax, 2010; Sürücü & Maslakçı, 2020).

2. Calculating the Average shared square variance (ASV) which is also a value between 1 and 0 and should be less than MSV to conclude that the discriminant validity is achieved (Sürücü & Maslakçı, 2020).

When a measurement tool is valid, it means that it measures one factor (variable, construct), converges with other tools that measure the same factor (when developing a new measurement tool, it is tested by comparing its results with other tools that measure the same variable), and support the measurement theory that is related to the measured factor (Heale & Twycross, 2015). On the other hand, validity is affected when using an inappropriate measurement tool, such as using complicated survey questions, or a long survey form with so many questions, giving limited time for the respondent to fill the survey form, and giving a short overall period for collecting the data (Thanasegaran, 2009). Generally, the more data that fits the measurement model is collected, the more validity is possible to be achieved (Schumacker & Lomax, 2010). After making sure that the data fits the scale used to measure the related variables and that there is reliability and validity, path analysis can be done for hypotheses testing.

4. Path Analysis and Its Related Assumptions

Path analysis, which is part of SEM, is an advanced version of multiple regression, used to examine hypothesized relationships between variables in much more complicated research models (Lleras, 2005; Streiner, 2005). These models may have more than one dependent variable or can represent a chain of a causal relationship between three or more variables where the first variable affects the second one then the second variable affects the third one (Lleras, 2005).

In other words, path analysis helps to examine the strength of direct and indirect relationships between the variables, where the first variable affects the second variable directly but indirectly affects the third variable due to the location of the second variable between them according to the hypothesized path (Lleras, 2005; Stage et al., 2004; Valenzuela & Bachmann, 2017;). This is a simple example of a path but paths can be more complicated with more variables and more relationships between them, the accumulation of all these direct and indirect relationships is called the variables' accumulated associations (Valenzuela & Bachmann, 2017).

Path analysis-related multivariate assumptions and how it is used for hypotheses testing are discussed in detail below:

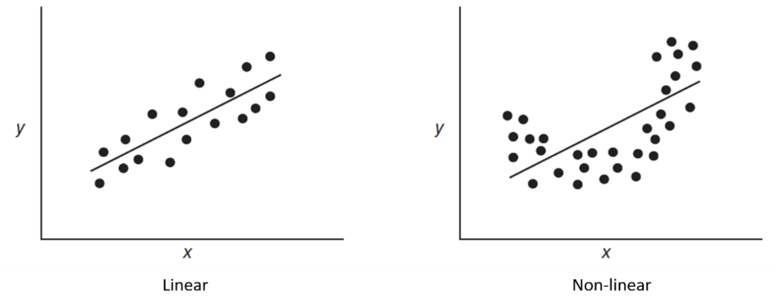
4.1. Path Analysis and Its Related Assumptions

Multivariate analyses represent statistical techniques that are used to analyze multiple measurements that measure a certain variable, they are also used to analyze research models that include more than one variable that explains, measures, and predicts the relationship between different variables in the research model (Hair Jr et al., 2014). Path analysis, which is an advanced multiple regression analysis, is one of the multivariate analyses. It is subjected to some multivariate assumptions such as normality, linearity, homoscedasticity, outliers, and multicollinearity.

- **Normality (Multivariate normality)** is a key assumption of most multivariate analyses (Oppong & Agbedra, 2016), it represents the probability symmetrical distribution of the data around their mean, where all the possible values of a variable are represented in the X horizontal axis, and the probabilities of each one of those values to happen are represented in the Y vertical axis (Hair Jr et al., 2014). It can be achieved more when having many observations, such as 200 observations or more (Hair Jr et al., 2014), as when having such a large number of observations it is directly assumed that the data distribution is approximately normal (Oppong & Agbedra, 2016). As explained earlier and shown in Figure 1. the distribution should not be skewed or kurtosised as this will negatively affect the variance-covariance between the variables (Schumacker & Lomax, 2010). It can be assessed by statistical graphs such as histograms (Oppong & Agbedra, 2016). In some cases, bootstrapping and normalization methods help in improving the data distribution to get a normal distribution (Boos, 2003).
- **Linearity** is a mathematical representation of the relationship between two correlated variables that is graphically demonstrated by a straight line (Hair Jr et al., 2014). It can be detected by using scatterplot graphs, as they help in evaluating the distribution pattern of the data, as shown in Figure 4. (Hair Jr et al., 2014; Schumacker & Lomax, 2010).

Figure 4

Linearity and nonlinearity

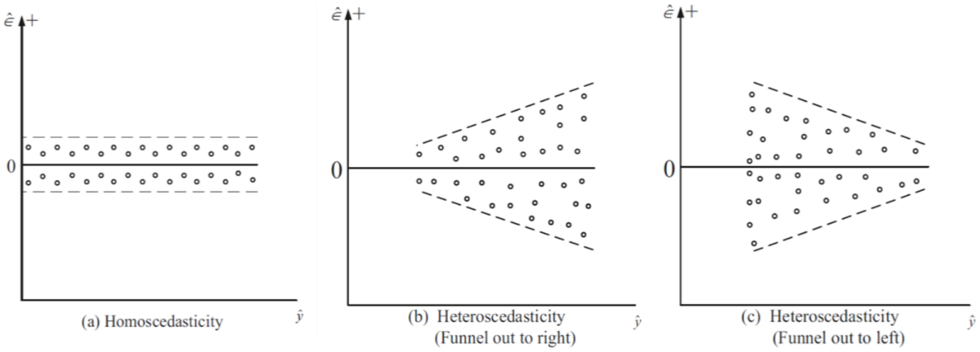


Source: (Montgomery et al., 2012).

- **Homoscedasticity** represents the variance of the dependent variable over the range of the independent variable which has (the variance) to be equal and constant, it can be calculated by the Levene test of variance, a Levene test's P-value which is more than 0.05 indicates that there is an equal constant variance. It also can be detected visually by evaluating the dispersion (variance) of the data when represented by scatter plots graphs as shown in Figure 5. (Hair Jr et al., 2014).

Figure 5

Homoscedasticity scatterplot.



Source: (Maiti, 2022).

- **Outliers** are observations that are extremely different from the remaining other observations; they have a unique value that is either very high or very low from the other observations' values (Hair Jr et al., 2014; Leys et al., 2019; Schumacker & Lomax, 2010). They can be detected by several methods such as scatter plots, box plots, frequency distributions, or histograms (Schumacker & Lomax, 2010). A high value of standard deviation is an indicator of the availability of outliers, however, the Z score value is used to find an outlier. Any observation with a standard Z score value of more than 3 is considered an outlier (Kline, 2011), for a larger sample size the standard Z score value cut point can be 4 (Hair Jr et al., 2014). Outlier observation has a negative impact on the calculation of the mean and standard deviation, as well as the correlation and regression analysis, so they have to be removed from the analysis (Schumacker & Lomax, 2010).

- **Multicollinearity** represents a linear relationship between the independent variable in a multiple regression model (Kumari, 2008; Shrestha, 2020), it occurs when the independent variables are significantly correlated with each other in addition to their correlation with the dependent variable (Shrestha, 2020). It leads to a high prediction error and causes significant variables to be insignificant (Kumari, 2008; Shrestha, 2020). It can be detected by calculating the variance inflation factor (VIF) which should be more than 10 to indicate a multicollinearity issue, it is also detected by calculating the Tolerance which should be between 0 and 0.1 to indicate a multicollinearity issue (Kumari, 2008; O'Brien, 2007). An independent variable with a high correlation with other independent variables can be removed to solve the multicollinearity issue, otherwise, additional data can be added to the current dataset or new data can be collected, then the model can be reassessed (Kumari, 2008).

Getting the path analysis assumptions is like getting the green light to go forward and start doing the path analysis and hypotheses testing.

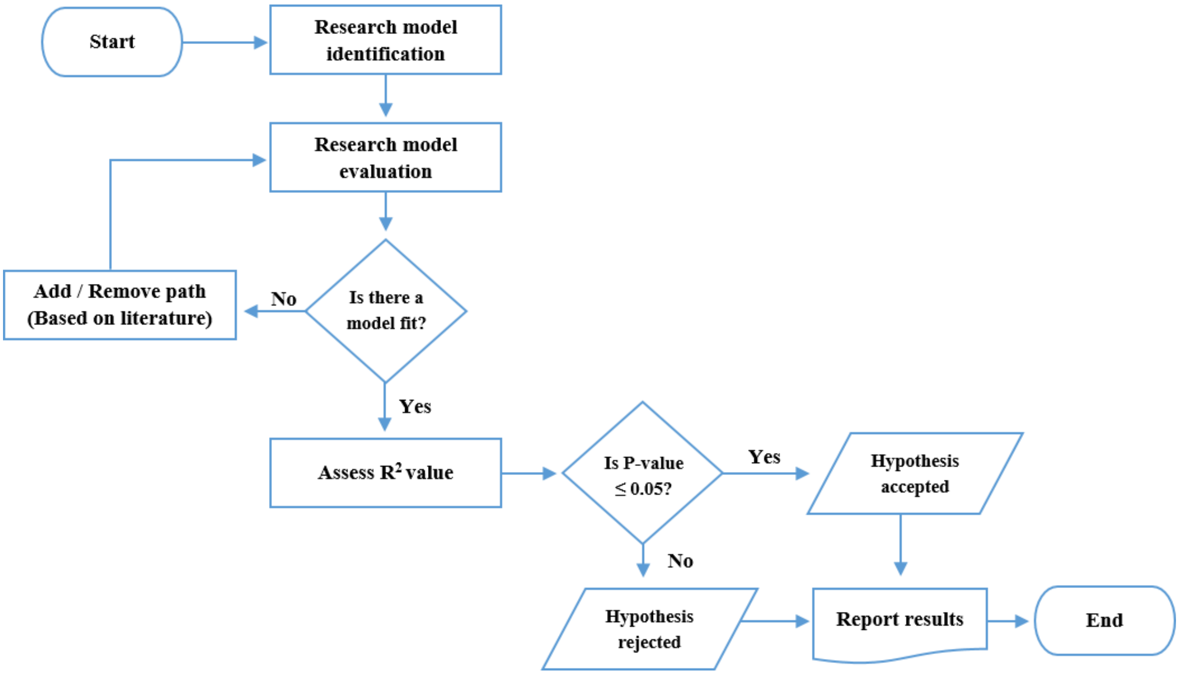
4.2. Path Analysis and Hypotheses Testing

Path analysis works on testing hypotheses by examining all the hypothesized relationships between the variables which are set according to a theoretical background, using a series of correlation and regression analyses (Valenzuela & Bachmann, 2017). The aim of a researcher when testing a hypothesis is to reject the null hypothesis H0, which states that there is no impact from the independent variable to the dependent variable (Hair Jr et al., 2014). There are steps to follow and implement before going directly to hypotheses testing, these steps are shown in Figure 6. and explained below:

- **Research model identification** is the first step in path analysis, in this step the researcher specifies which variables to include in the research model and defines the relationships as well as the direction of influence between them, this is done by having a solid theoretical foundation of the research model, based on a theoretical background and previous literature (Streiner, 2005; Valenzuela & Bachmann, 2017). It is not recommended to include as much as variables or paths in the research model and try if things will work out, as path analysis aims to test not build a research model (Streiner, 2005), Adding unrelated variables or paths or even removing a variable or a path without having a theoretical base to do so, leads to changing the analysis results due to changing the research model (Lleras, 2005).

Figure 6

Path Analysis Steps.



Source: Developed by the author, based on (Hair Jr et al., 2014; Miyao & Maruyama, 2012).

In this step, variables are divided into exogenous variables and endogenous variables. The exogenous variables are not affected by any other variables in the model but can affect them such as the independent variables. On the other hand, the endogenous variables are affected by other variables in the model, some of them can also affect other variables in the model such as the mediating variables, but others can not affect them such as the dependent variable. A moderating variable can be either exogenous or endogenous, based on whether it receives an effect from any other variable in the model or not (Lleras, 2005; Valenzuela & Bachmann, 2017).

Finally, in this step using analysis software such as AMOS and LISREL the research model can be presented visually, each variable is represented by a rectangular shape, variables are associated with single-head arrows to show the path direction and how the effect flows from exogenous to endogenous variables, double-head arrows are used to link between the exogenous variables (to see the correlation between them which has to be as less as possible), and every endogenous variable should be linked with an error term (Lleras, 2005; Valenzuela & Bachmann, 2017).

- **Research model evaluation (model fit)** comes directly after identifying the research model in path analysis compared to CFA, there is no factor loading to estimate. The research model fit is evaluated to make sure that the collected data fit the hypothesized research model, the same method and same fitting indices used in CFA are also used in path analysis (Lleras, 2005; Streiner, 2005).

  Model fit evaluation helps the researcher to compare between research models which one of them fits the data more, especially when having several theories that represent the variables under the study and the paths associating them, it also helps to evaluate the research model fit when a combined research model is developed based on the combination of two or more theories (Lleras, 2005).

  The research model fit could be lost when the research model includes some variables or paths that are not related to each other or are not theoretically justified, this can happen when a path is added between two variables that are not related to each other, having exogenous variables that are not related to any endogenous variables, or removing a critical path or variable (Streiner, 2005).

  Sample size plays a role in the research model fit evaluation as the more complicated the model the more samples it requires. In some literature it is recommended to multiply the number of parameters (factors, structures, variables) by 20 to have an adequate amount of data to test the research model, some other literature recommends having at least 250 to 300 samples to test the research model (Stage et al., 2004; Valenzuela & Bachmann, 2017), however, having more sample is always better as it leads to the decrease of the margin of error and the increase of the power the research (Reyes & Ghosh, 2013).

- **R-squared (R^2)** is a fit value that shows how the regression results fit the model, it represents the amount of variance in the dependent variable caused by the independent variables, in other words, it represents the amount of impact from the independent to the dependent variables (Byrne, 2010; Streiner, 2005). It is a value between 1 and 0, the higher the percentage the better the results, as a low R^2 value is an indicator of a high margin of error (Byrne, 2010; Hooper et al., 2008). It is affected by the complexity of the research model and the sample size, having an R^2 value less than 0.20 shows that there is an impact from the independent variable to the dependent variable, but this impact is low due to of high margin of error (Hooper et al., 2008).

- **Research model amendments:** Path analysis is sensitive to the identification of the research model, as any changes in the model such as adding irrelevant variables or removing relevant ones can affect the results (Streiner, 2005). However, when the research model doesn't fit the collected data or when the R^2 value needs to be improved, the researcher can add, remove, or change the direction of the paths in the model, but based on a solid theoretical foundation (Valenzuela & Bachmann, 2017).

  Researchers can also remove variables based on a solid theoretical foundation to improve the model fit and R^2, but in such cases, factor analysis, as well as reliability and validity assessment, should be done again because removing a variable means that the research model is changed and analysis should be updated to be consistent with the new model (Valenzuela & Bachmann, 2017).

  Each element of the research model should make sense, adjusting the model randomly to improve the model fit and R^2 leads to irrelevant results (Streiner, 2005). Recollecting the data can also help in improving the model fit and R^2, especially if the appropriate sample is targeted.

- **Hypotheses testing** comes after making sure the collected data fits the research model and an appropriate R^2 is acquired. The probability P-value is calculated in this step to see whether to accept a hypothesis or not (Carvalho & Chima, 2014), a P-value of 0.05 or less is an indicator of accepting a hypothesis (Hair Jr et al., 2014).

  When conducting path analysis for hypotheses testing standardized and unstandardized paths are calculated, however, the standardized paths are used when reporting the results (Valenzuela & Bachmann, 2017).

When making a comparison path analysis's key strength point over multiple regression that attracts the researchers is its ability to evaluate the goodness of fit between the research model and the collected data, as well as its ability to test the direct and indirect relationships between variables in complicated research models (Stage et al., 2004; Valenzuela & Bachmann, 2017).

5. Conclusion

It is concluded that SEM is an advanced multivariate analysis, that includes a series of steps that help in testing complicated research models for social and behavioral studies. Such models include several constructs or factors, that reflect a group of hypotheses that link between independent and dependent variables. These variables are set in a certain order that needs a series of multiple regression analyses to explain the nature of relationships between them. In SEM, as same as any other analysis, before going directly to hypotheses testing there are several procedures to be done starting by examining the quality of the collected data by conducting data screening tests and descriptive statistics. Then the scale (measurement theory) that is used to measure the research variable should be evaluated by conducting factor analysis (exploratory then confirmatory), which helps to make sure that the collected data fits the measurement model as well as assess the scale’s reliability and validity. After that, a quick review of the multivariate assumptions should be done, which leads finally, to path analysis and hypotheses testing. The researcher evaluation is important to make the final decision of accepting each step of the analysis outputs and values and moving forward to the next step. The accumulation of these steps with their appropriate sequence, which is explained above, leads to a well-conducted SEM analysis for social and behavioral studies, with results that can be explained and interpreted based on the theoretical background and literature foundation of the research.

6. Limitations and Further Reading Recommendations

This paper explained the steps of conducting SEM, for quantitative analysis students and other researchers who have basic knowledge of statistics, starting with evaluating the data and evaluating the scale then hypotheses testing. However, it didn’t discuss the details of statistical techniques used to test each type of relationship between variables, such as mediation and moderation. It also didn’t discuss how to deal with control variables or group comparisons. So, it is recommended after reading this paper to read and search about these topics to move forward in getting knowledge about statistics and hypotheses testing in social and behavioral science.

Thanks: To Dr. Güven Ordun – Istanbul University for his help in developing Figure 2.

References

Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020). Chi-square for model fit in confirmatory factor analysis. *Journal of Advanced Nursing, 76*, 2209-2211. https://doi.org/10.1111/jan.14399

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, *18*, 6. https://doi.org/10.7275/qv2q-rk76

Bonett, D. G., & Wright, T. A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior, 36*(1), 3-15. https://doi.org/10.1002/job.1960

Boos, D. D. (2003). Introduction to the bootstrap world. *Statistical Science, 18*(2), 168–174. https://doi.org/10.1214/ss/1063994971

Brown, J. D. (2002). The cronbach alpha reliability estimate. *JALT Testing & Evaluation SIG Newsletter, 6*(1), 17-18. http://jalt.org/test/bro_13.htm

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2 ed.). The Guilford Press.

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

Byrne, B. M. (2010). *Structural equation modeling with amos basic concepts, applications, and programming* (2nd ed.). Taylor and Francis Group, LLC.

Carvalho, J., & Chima, F. O. (2014). Applications of structural equation modeling in social sciences research. *American International Journal of Contemporary Research*, *4*(1), 6-11. https://doi.org/10.30845/aijcr

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250-1262. https://doi.org/10.1016/j.jbusres.2008.01.013

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4-19. https://doi.org/10.1016/j.jesp.2015.07.006

DeSimone, J. A., Harms, P. D., & Desimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, *36*(2), 171-181. https://doi.org/10.1002/job.1962

Ercan, I., Yazici, B., Sigirli, D., Ediz, B., & Kan, I. (2007). Examining cronbach alpha, theta, omega reliability coefficients according to the sample size. *Journal of Modern Applied Statistical Methods*, *6*(1), 291-303. https://doi.org/10.22237/jmasm/1177993560

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications Ltd.

Fisher, M. J., & Marshall, A. P. (2009). Understanding descriptive statistics. *Australian Critical Care, 22*(2), 93-97. https://doi.org/10.1016/j.aucc.2008.11.003

Fitzner, K. (2007). Reliability and validity a quick review. *The Diabetes Educator*, *33*(5), 775-780. https://doi.org/10.1177/0145721707308172

Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using pls-graph: Tutorial and annotated example. *Communications of the Association for Information Systems*, *16*(5), 91-109. https://aisel.aisnet.org/cais/vol16/iss1/5?utm_source=aisel.aisnet.org%2Fcais%2Fvol16%2Fiss1%2F5&utm_medium=PDF&utm_campaign=PDFCoverPages

Griffin, M. M., & Steinbrecher, T. D. (2013). Large-scale datasets in special education research. In R. M. Hodapp, D. J. Fidler, R. M. Hodapp, & D. J. Fidler (Eds.), *International review of research in developmental disabilities* (pp. 155-183). Academic Press. https://doi.org/10.1016/B978-0-12-407760-7.00004-9

Hair, J. F., Hult, G. T., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3 ed.). Sage.

Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson Education Limited.

Hamid, M. A., Sami, W., & Sidek, M. M. (2017). Discriminant validity assessment: Use of Fornell & Larcker criterion versus HTMT criterion. *In Journal of Physics: Conference Series*, *890*(1), 012163. https://doi.org/10.1088/1742-6596/890/1/012163

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, *18*(3), 66-67. https://doi.org/10.1136/eb-2015-102129

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393-416. https://doi.org/10.1177/0013164405282485

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53-60.

Hu, L.T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *5*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2015). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology*, *30*, 299–311. https://doi.org/10.1007/s10869-014-9357-6

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, *27*(1), 99-114. https://doi.org/10.1007/s10869-011-9231-8

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology, 100*(3), 828–845. https://psycnet.apa.org/doi/10.1037/a0038510

Jhangiani, R. S., Chiang, I.-C. A., Cuttler, C. & Leighton, D. C. (2019). *Research methods in psychology* (4th ed.) Kwantlen Polytechnic University.

Jöreskog, K.G., & Sörbom, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research*, *19*(4), 404-416. https://doi.org/10.1177/002224378201900402

Kenny, D. A. (2020). *Measuring model fit*. Retrieved September 4, 2022, from https://davidakenny.net/cm/fit.htm

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). The Guilford Press.

Kumari, S. S. (2008). Multicollinearity: Estimation and elimination. *Journal of Contemporary Research in Management, 3*(1), 87-95.

Lee, D. K., In, J., & Lee, S. (2015). Standard deviation and standard error of the mean. *Korean Journal of Anesthesiology*, *68*(3), 220-223. https://doi.org/10.4097/kjae.2015.68.3.220

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology, 32*(1), 1–10. https://doi.org/10.5334/irsp.289

Livingston, E. H. (2004). The mean and standard deviation: What does it all mean? *Journal of Surgical Research*, *119*(2), 117–123. https://doi.org/10.1016/j.jss.2004.02.008

Lleras, C. (2005). Path analysis. *Encyclopedia of Social Measurement, 3*(1), 25-30.

Maiti, J. (2022). *Multivariate statistical modeling in engineering and management* (1st ed.). Taylor & Francis Group, LLC. https://doi.org/10.1201/9781003303060

Marshall, G., & Jonker, L. (2010). An introduction to descriptive statistics: A review and practical guide. *Radiography*, *16*(4), e1-e7. https://doi.org/10.1016/j.radi.2010.01.001

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437-455. https://psycnet.apa.org/doi/10.1037/a0028085

Miyao, H., & Maruyama, R. (2012). On-line handwritten flowchart recognition, beautification, and editing system. *International Conference on Frontiers in Handwriting Recognition* (pp. 83-88). IEEE. https://doi.org/10.1109/ICFHR.2012.250

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). John Wiley & Sons, Inc.

Muijs, D. (2004). *Doing quantitative research in education with SPSS* (1st ed.). Sage Publications. http://digital.casalini.it/9781526472779

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*, 673–690. https://doi.org/10.1007/s11135-006-9018-6

Oppong, F. B., & Agbedra, S. Y. (2016). Assessing univariate and multivariate normality, a guide for non-statisticians. *Mathematical Theory and Modeling, 6*(2), 26-33.

Orcan, F. (2018). Exploratory and confirmatory factor analysis: Which one to use first? *Journal of Measurement and Evaluation in Education and Psychology, 9*(4), 414-421. https://doi.org/10.21031/epod.394323

Reyes, E. M., & Ghosh, S. K. (2013). Bayesian average error-based approach to sample size calculations for hypothesis testing. *Journal of Biopharmaceutical Statistics, 23*(3), 569-588. https://doi.org/10.1080/10543406.2012.755994

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). Taylor and Francis Group, LLC.

Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics, 8*(2), 39-42. https://doi.org/10.12691/ajams-8-2-1

Stage, F. K., Carter, H. C., & Nora, A. (2004). Path analysis: An introduction and analysis of a decade of research. *The Journal of Educational Research, 98*(1), 5-13. https://doi.org/10.3200/JOER.98.1.5-13

Streiner, D. L. (2005). Finding our way: An introduction to path analysis. *The Canadian Journal of Psychiatry, 50*(2), 15-122. https://doi.org/10.1177/070674370505000207

Suhr, D. D. (2006). Exploratory or confirmatory factor analysis? CA: SUGI 31.

Sürücü, L., & Maslakçı, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal, 8*(3), 2694-2726. https://doi.org/10.15295/bmij.v8i3.1540

Tabachnick, B. G. & Fidell, L. S., 2013. *Using multivariate statistics* (6th ed.) Pearson.

Tavakol, M., & Dennick, R. (2011). Making sense of cronbach's alpha. *International Journal of Medical Education, 2*, 53-55. https://doi.org/10.5116%2Fijme.4dfb.8dfd

Thanasegaran, G. (2009). Reliability and validity issues in research. *Integration & Dissemination, 4*, 35-40.

Thompson, C. B. (2009). Descriptive data analysis. *Air Medical Journal, 28*(2), 56-59. https://doi.org/10.1016/j.amj.2008.12.001

Valenzuela, S., & Bachmann, I. (2017). Path analysis. *The International Encyclopedia of Communication Research Methods*, 1-9.

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking internal consistency in cronbach's alpha. *Leisure Sciences, 39*(2), 163-173. https://doi.org/10.1080/01490400.2015.1127189

Weston, R. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist, 34*(5), 719-751. https://doi.org/10.1177/0011000006286345

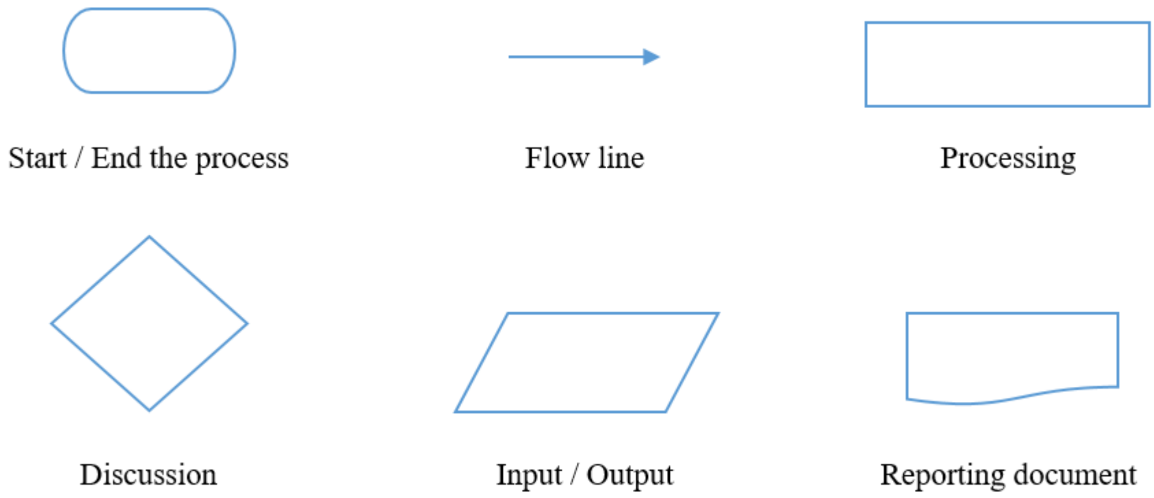
Yaşlıoğlu, M., & Toplu Yaşlıoğlu, D. (2020). How and when to use which fit indices? A practical and critical review of the methodology. *Istanbul Management Journal, 88*(1), 1-20. https://doi.org/10.26650/imj.2020.88.0001

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79-94. https://doi.org/10.20982/tqmp.09.2.p079

Journal of Organizational Behavior Review (JOBReview)
Cilt/Vol.: 6 Sayı/Is.: 1 Yıl/Year: 2024 Sayfa/Pages: 69-108

Appendix: Flow Charts Shapes Explanation

Below is an explanation of the flowchart shapes presented in Figure 3. and Figure 6.



Source: (Miyao & Maruyama, 2012).