Journal of Management Sciences

Journal of Management Sciences

Manualizing Factor Analysis of Likert Scale Data

Affiliation:

Nadeem Uz Zaman
Assistant Professor, Department of Management Sciences, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta. Email: nadeem.zaman@buitms.edu.pk

Zainab Bibi
Chairperson/ Professor, Institute of Management Sciences, University of Balochistan, Quetta. Email: zainab.ims@uob.edu.pk

Sana Ur Rehman Sheikh
Assistant Professor, Department of Business Administration, NFC Institute of Engineering and Technology, Multan, Pakistan. Email: dr.sana.ur.rehman@nfciet.edu.pk

Abdul Raziq
Dean/ Professor, Faculty of Management Sciences, University of Loralai, Pakistan. Email: dr.abdulraziq@uoli.edu.pk

Manuscript Information
Submission Date: August 22, 2020
Acceptance Date: September 18, 2020
Publication Date: September 26, 2020

Citation in APA Style:
Zaman, N., Bibi, Z., Sheikh, S. R., & Raziq, A. (2020). Manualizing Factor Analysis of Likert Scale Data, *Journal of Management Sciences, 7*(2), 56-67.

DOI: https://doi.org/10.20547/jms.2014.2007204

.

Journal of Management Sciences
Vol: 7(2): 56-67, 2020
DOI: 10.20547/jms.2014.2007204

Manualizing Factor Analysis of Likert Scale Data

Nadeem Uz Zaman [*] Zainab Bibi [†] Sana Ur Rehman Sheikh [‡]
Abdul Raziq [§]

Abstract: *Despite the academic value of factor analysis (FA) on Likert scale data, its statistical legality has been relentlessly questioned by professional statisticians. This article reviews relevant literature and proposes a statistically appropriate method of getting most out of FA on Likert scale. The review shows that a larger sample size is an important consideration in improving the appropriateness of FA on LS as it improves the solution in terms of normality, communalities and loadings. Further, a 7-9 point or greater scale to account for normality, an alpha level of 0.01 or 0.005, polychoric correlation instead of Pearson's are reported to improve the statistical appropriateness of the test of Likert scale. Moreover, using non-parametric alternatives like CATPCA to testify the results of FA greatly increased the overall value and validity of the test.*

Keywords: Likert scale, factor analysis, validity, multivariate data analysis.

Introduction

Likert scale was first introduced by Likert (1932) as a measure of attitude or opinion on an odd-numbered response set with options including 'strongly approve', 'somewhat approve', 'no idea', 'somewhat disapprove' and 'strongly disapprove'. The scale was later used in many diverse variations in academic studies and business research including measurements of happiness, intelligence, completeness, excellence, dullness, superiority, priority, importance and so on (Clason & Dormody, 1994; Vogt, 1999). In truth, the usefulness of the scale is almost unanimously signified across various disciplines (Balasubramanian, 2012; Barua et al., 2013; Clason & Dormody, 1994). Veritably, looking at the generous use of the scale in myriads of studies, one can easily infer that without LS much of modern-day academic understanding would have been merely a mirage.

Despite the value of LS in our expansion of knowledge, the scale itself is debated for the appropriateness of statistical analyses. Ironically, it has become somewhat commonplace to notice that many researchers misuse LS as a very pliable measurement scale

[*] Assistant Professor, Department of Management Sciences, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta. Email: nadeem.zaman@buitms.edu.pk

[†] Chairperson/ Professor, Institute of Management Sciences, University of Balochistan, Quetta. Email: zainab.ims@uob.edu.pk

[‡] Assistant Professor, Department of Business Administration, NFC Institute of Engineering and Technology, Multan, Pakistan. Email: dr.sana.ur.rehman@nfciet.edu.pk

[§] Dean/ Professor, Faculty of Management Sciences, University of Loralai, Pakistan. Email: dr.abdulraziq@uoli.edu.pk

Received 22 August 2020; Received in revised form 10 September 2020; Accepted 18 September 2020 Available online 26 September 2020

requiring no good knowledge (Allen & Seaman, 2007; Bishop & Herron, 2015; Boone & Boone, 2012; Carifio & Perla, 2007). This thinking causes several technical flaws when it comes to the statistical analyses of LS. One of such cases can be found in the use of factor analysis (FA), one of the most widely used statistical techniques on LS data often to test the validity of data collection instrument (Marascuilo & Levin, 1983).

Even though there are many literary works that discuss some aspects of FA on LS data, we did not come across a comprehensive composition about the most appropriate method of employing this technique. A general reading shows that the literature relevant to the question is scattered in pieces to be bundled together for a convenient usage. At the same time, there are some general suggestions in some works on FA which could well be associated with LS data so as to improve the validity and usage of the solution. Thus, in this paper we reviewed numerous relevant statistical literatures to suggest how to get most out of the FA on LS data. We, nonetheless, assumed that the reader has some prior knowledge of FA. Secondarily, as there is sufficient reasoning in the existing studies about the suggestions placed in this paper, we resorted merely to the presentation of ideas rather than arguments. The suggestions themselves are organized into simple, manageable headings.

An overview of the use of FA on LS data indicates that researchers are often not able to differentiate between a 'scale' and a 'response item' particularly in non-statistical fields of study (Carifio & Perla, 2007; Boone & Boone, 2012; Brown, 2011; Jamieson, 2004). Most often it is an autonomous and independent Likert type question (Likert scale item) that a researcher is trying to analyze rather than data on LS (the scaled instrument) which is a reasoned and a structured whole. A LS composes of many items, at least 3 to 5- measuring some single attitude using a set of questions (Wheelwright, Levine, Garfinkel-Castro, Bushman, & Brewer, 2020; Boone & Boone, 2012). Whereas questionnaires with only one item or question measuring an aspect of some attitude does by no means become a scale: it is just a question. In order to uncover the latent structure among variables, LS is required as a response format with a set of items (Lubke & Muthén, 2004; Warmbrod, 2014).

Factor analysis is a very popular statistical method used for determining the structure among certain manifest variables in terms of the underlying latent variables (KILIÇ, n.d.). Researchers across diverse disciplines have been using FA in several ways including testing the validity of research instrument and theory (Garson, 2010; Pett, Lackey, & Sullivan, 2003; Tabachnick & Fidell, 2001). It is, however, important to note that in order to improve the quality of FA, researcher have to make some subjective decisions and choices. It is the quality of these decisions on which the accuracy of FA rests (Henson & Roberts, 2006; Tabachnick & Fidell, 2001). Furthermore, often in applying FA, a research has to go through several cyclical refining of the results so as to reach an acceptable solution. This makes FA more complex than is commonly thought.

Unfortunately, the widespread use of FA on Likert Scale has made researchers think that it is perfectly valid to use FA on Likert Scale as it is used on any type of data (Brown, 2011; Wheelwright et al., 2020). It is even more unfortunate that even academic literature on LS itself is not spared of misunderstandings and misconceptions. In fact, there are certain circumstances that render FA altogether erroneous on LS and results in misleading generalizations (Clason & Dormody, 1994). The problem is that despite being coded into numbers, a Likert scale is in fact a set of ordered categories (Brown, 2011) with unequal scale values. We, in response, try to put forth some suggestions to improve the legality and validity of FA on LS.

Is FA an Appropriate Technique for LS?

With a very strict view, the particular structure of LS might look like an ordinal scale which results in specific restrictions on its statistical analyses (Gorsuch, 1990; Knapp, 1990). It is this reason that some researchers believe that parametric statistics are not appropriate for LS (Vigderhous, 1977; Jakobsson, 2004; Jamieson, 2004). In order to run FA as a parametric technique, data must meet strict assumptions of continuity, linearity, absence of extreme multicollinearity, absence of outliers and low percentage of missing values (Pett et al., 2003). Moreover, whether or not LS returns continuous data, is another debatable question.

Literature on LS being continuous or not is divided into debatable conclusions. One stance is that the scale is not continuous at all in pure statistical terminology (Clason & Dormody, 1994). However, there is general tendency of researchers to presume a latent or natural continuity in the scale, which, if true, makes the scale resemble at most to a sort-of 'quasi-interval scale' only just for the sake of statistical adventurism (Brown, 2011; Clason & Dormody, 1994). This is not all that makes LS become quasi-interval with some latent continuity; in order for it to be so the population from which the data was drawn must be fairly ordered- or strictly speaking, homogenous across all latent variables, which does not often take place. However, in order to consider such a quasi-interval scale, it must pass Cronbach's alpha or Kappa test of inter-correlation and validity for consideration of FA.

On the contrary, some statisticians advocate that the very definition of scales of measurement itself is arbitrary (Angoff, 1984; Knapp, 1990). Therefore, the requirement of data continuity is rather overemphasized and factor analysis, or any other parametric tests can be used on LS data. But at the same time some assumptions should be entertained for valid results including multivariate normality of data and a given number of categories (Lubke & Muthén, 2004). These claims have been supported by empirical evidence in some studies as accurate test statistics of LS data have been reported using parametric analyses. Nevertheless, some experts suggest that increasing the number of response categories can, to some extent, improve the level of 'quasi-normality' in LS. For example, Knapp (1990) suggests that a 10-point LS is more continuous or less discrete than a 5-point scale.

We believe that FA is very useful in adding meaningful value to research and should be allowed to use in analyzing LS data (Allen & Seaman, 2007). Nonetheless, it is important that the tests be unbiased and convincing. As a matter of fact, despite strict call for compliance with scales of measurement and statistical tests there remains an extravaganza of discords on the very definitions as to when a scale is ordinal or interval as well as its usage (Knapp, 1990). In exact, the differences are mainly due to theoretical perspective. However, the question of data normality remains (Thomas, 1982; Wheelwright et al.,

2020) as an important ingredient of FA.

FA and Principal Component Analysis PCA

One of most widespread practices in running FA is using Principle Component Analysis (PCA) in some popular software designed for social sciences- SPSS and SAS for example. However, there is disagreement among statisticians if PCA is a true alternative of FA and there are arguments both in favor and against it (Gorsuch, 2003; Mulaik, 2009; Schreiber, 2020). Gorsuch (2003) suggests that PCA has become famous because it was commonly used when computers were too slow for FA.

PCA differs from FA in several stances. At first unlike FA, PCA disregards the underlying structure of the latent variables and components are extracted through variance within the manifest variables (Fabozzi, Focardi, Rachev, & Arshanapalli, 2014; Suhr, 2006; Zöller, 2012). Moreover, in FA the shared variance of variables is partitioned into unique variance and error variance and during the factor extraction; PCA does not make such discriminations which results in overestimated variances for the components. This paper, consequently, focuses only on FA disregarding PCA in any consequent discussion. Any further reference to PCA is thus not made and the paper exclusively talks about FA only.

Normality of Data

FA assumes that data are normally distributed albeit not a quintessential requirement (Tabachnick & Fidell, 2001; Wheelwright et al., 2020). Nevertheless, it is generally important for the validity of any parametric tests including FA that depend upon the extent to which the data are normally distributed. Data can be assumed to be normally distributed if the sample size is fairly large in accordance with the central limit theorem. This can be withheld without necessarily resorting to any statistical tests or even if the test statistics point to non-normality of the data (Lumley, Diehr, Emerson, & Chen, 2002). The downside is that the tests used for normality are reported for not being strong enough to accurately measure normality for every type of data (Doane & Seward, 2011; Lumley et al., 2002). Since LS is not a continuous scale in its true sense, normality of data on LS might be assumed with a sense of theoretical normality for which a fairly large sample size can be suggested to account for normality of data collected on LS. Normality of data, consequently, ensures that mean, median and mode are the same for a population under study (Anderson, Sweeney, Williams, Camm, & Cochran, 2020; Utts, 2014). We will discuss more about the sample size consideration on LS in the latter sections.

Correlation

Correlations among the variables is an essential determinant of the validity of FA. A minimum of .30 correlation coefficient is suggested to decide if one can run FA (Tabachnick & Fidell, 2001). Likert scale present a special case when it comes to estimating correlation and hence requires different treatment. The common practice regarding FA on LS is the familiar use of Pearson correlation, mostly as a default option in several software. Since data on LS are not completely continuous, Pearson correlation cannot be used as a measure of association as this test is sensitive to scale responses and using it is misleading (Clason & Dormody, 1994).

Though in ordinal scales with binary responses Spearman correlation is preferable to Pearson correlation, in case of polytomous data as is the case with LS, polychoric correlation is more advisable (Byrne, 2006). The same is also true for maximum likelihood estimation on LS. In order to go around the issue of default options in the software like SPSS or SAS, custom-build software with polychoric correlation can be used. FACTOR is one of such software that are freely available on the internet. Alternatively, polychoric correlation can be calculated using software like R and used to run FA on it.

Extraction of Factors

Under common practices, Kaiser Criterion of eigenvalue is implied to extract components of a factor. Under this criterion any factor with eigenvalue greater than 1 is retained in the final solution. However, researchers suggest that this is an inaccurate method that often results in too many factors being retained- a situation called overdetermination (Auerswald & Moshagen, 2019; Goretzko, Pham, & Bühner, 2019). This is because, often the eigenvalue criterion (Kaiser Criterion) does not result in the actual factors that exist within the patterns of data. In order to see if there is overdetermination in the solution several suggestions have been made in the literature (Schreiber, 2020; Sellbom & Tellegen, 2019). Overdetermination can be traced through factor-to-variable ratio. On the other hand, many items in a factor reflect overdetermination if the communalities fall below 0.50. Whereas, a minimum of 3 items per factor as critical benchmark; however, Costello and Osborne (2005) claim that any factor with 3 items is unstable and weak. Likewise, Fabrigar et al, (1999) are of the opinion that at least four items should comprise a factor up to as many as six items.

Instead of using eigenvalue greater than '1' as the criteria to retain factors, (Costello & Osborne, 2005) suggest that the use of scree plot is more appropriate. It is a subjective and manual method of extracting maximum number of factors using scree plot to look for the natural break 'flattening of the curve' instead of eigenvalues. So, when the curve starts becoming flat, that point can be considered as the cut-point to extract maximum components. However, in some cases several FA should be run to see if the factor retained correspond to factors expected. Nonetheless, if the FA fails to correspond to theory- expected numbers of factors- then there must be some problem with the data itself (Schreiber, 2020). Though this rule is a general one, it is very useful in case of LS because the scale does not meet the strict assumption of FA. So, relying on eigenvalue to correspond to the theoretical underpinning of the latent variables might not a prudent practice (Goretzko et al., 2019).

Rotation

The choice of rotation method is among the most important issues with FA on LS data. Under common practice, researcher often tend to use Varimax as the method of rotation across disciplines. Varimax and other orthogonal rotations like quartimax and equimax result in factors that are mutually uncorrelated (Goretzko et al., 2019). Oblique rotations, on the other hand, return factors that are correlated. Some of the most commonly used oblique rotations include direct oblimin, quartimin and promax. As opposed to what is commonly practiced, (Beavers, Lounsbury, Richards, & Huck, 2013; Costello & Osborne, 2005) suggest that in social sciences orthogonal rotations are not appropriate despite the conventional use. They argue that in social science, researcher should expect that some correlation among factors does exist and they should provide for such correlations in the analyses. Thus, it is rather appropriate to use oblique rotation methods. Further, it is quite interesting to note that using an oblique rotation is a rather logical choice because even if the factors are not correlated at all, oblique rotation will result in a zero-correlation coefficient for components (Schreiber, 2020). Orthogonal rotations, on the other hand, do not provide for such a built-in adjustment.

Keeping in view the point raised by Beavers et al. (2013); Costello and Osborne (2005), we accordingly propose that as LS is a commonly employed scale of measurement in social sciences, it is quite appropriate to use oblique rotation instead of orthogonal rotation methods as long as there is sufficient evidence otherwise to suggest that factor should not be correlated. In case one decides to use orthogonal rotation anyhow; Fabrigar et al. (1999) suggest that any oblique rotation method including varimax, quartimax and equimax produces more or less that same results. Therefore, no priority can be set for any particular methods of rotation. They also advocate that it is rather appropriate to use the default settings of the software to avoid unnecessary complications.

Sample Size

The most important question in establishing validity of FA on LS is that of sample size. In order to meet the assumptions required for FA, several recommendations are made in the literature as to what sample size suffices the appropriateness of FA; for example as a sample size of 50 (Lawley & Maxwell, 1971), 150, 100 (Gorsuch, 1990; Kline, 1979), 200, 250, 300 and so on. Furthermore, Lawley and Maxwell (1971) have suggested that sample size should be at least 51 cases more than the items in the instrument.

Alternatively, subject-to-variables (STV) ratio can also be used to determine a sample size appropriate for FA (Goretzko et al., 2019). Several suggestion in this regard have also been put forth as to what is the appropriate STV including 20:1, 10:1, not lower than 5:1, 3:1 to 6:1 but if sample size is at least 250 and at least 2:1 with a minimum of 100 subjects (Kline, 1979). The implication of sample size greatly affects the validity of FA solution (Hogarty, Hines, Kromrey, Ferron, & Mumford, 2005; Lawley & Maxwell, 1971).

Despite the emphasis placed on larger sample sizes to ensure the validity of FA on LS, it should also noticed that sample size itself has a diminishing marginal improvement in the validity of FA solution and thus excessively large sample size are also not desirable: it is like too much of a good thing (Bacchetti, Deeks, & McCune, 2011; Dolnicar, Grün, & Leisch, 2016). What implies is this that one should just suffice a data size that appropriately justifies the use of FA rather than insisting on wasting time and resources in collecting useless data.

The question of sample size is of relevance to justify data normality, low factor loadings and communalities in FA solution on LS. Sufficiently large sample size ensures that the normality of data is maintained in line with the central limit theorem (Anderson et al., 2020; Lumley et al., 2002). Secondly, larger sample size- exceeding 300- provide sufficient legality for FA solutions when there are only few highly correlated variables and when it is required to collapse highly multicollinear cases. Thirdly, to provide justification of FA when factor loadings fall below 0.4. It is also proposed that a sample size of at least 300 is indispensable. The same is the case for solutions with low communalities and few variables loading on each factor. In such a case sample size exceeding 100 subjects can provide some justification for FA.

Loadings

Loadings account for the unique variance of a variable that explains a factor and define the structure of different factors. The factor loadings of each variable are linear combinations that mathematically summarize the relationships among variables and the factors. They are strong predictors of congruence between sample and population (Costello & Osborne, 2005). Thus, the sample-population pattern fit is very good for values of loadings greater than 0.80, acceptable for value equal to 0.60 and very poor for 0.40. A latent variable with 5 or more items with loadings greater than 0.5 is a very strong factor. However, if the loadings fall as low as 0.4, the solution should only be interpreted if the sample size is greater than 300 cases since, as the sample size increases, the standard error related with factor loading tends to decrease.

Communalities

Sample size also has elemental relationship with communalities. Communalities are the values representing the unique variance of a variables that is finally explained after the variable has become part of a factor. This concept is of importance as when a variable is included in a factor, it loses some of its shared variance and only a part of it is expressed in the factor (Goretzko et al., 2019). Costello and Osborne (2005) suggest that in social sciences it is difficult to achieve communalities that are high (0.8 of greater), they rather fall somewhere between 0.4 and 0.7 in most cases. To go around this limitation in social sciences, (Tabachnick & Fidell, 2001) advocate that this problem can be handled with sufficiently large sample sizes. So, 0.40 is an acceptable value for communalities in case of LS, provided the sample size is fairly large. Similarly, Tabachnick and Fidell (2001) recommend that communalities even as low as 0.3 are enough evidence to justify FA if the sample size is large.

We conclude our discussion about an appropriate sample size that provides for factor loading, communalities as well as normality and adequacy of data in the overall FA solution on LS data. Keeping in view the marginal improvement in the data properties we need to see that we do not waste our resources and time in uselessly endeavoring to collect excessive large samples of data. Therefore, in the light of our review we pick a mediocre value of sample size which is somewhere equal to 300 cases and we suggest that this sample size should be sufficient to ensure that data are normally distributed, the researcher can even interpret the FA solution with factor loadings and communalities falling as low as 0.4 and 0.3 respectively as it is often the case in social sciences.

Some other Suggestions

If intercorrelations are unexpectedly low, it may be a result of low variance due to high level of homogeneity in the sample. Samples that are too homogenous can exhibit low variance; consequently, the correlation will be low potentially failing to reveal a factor, or common relationship, that does exist. Use a p-value as low as 0.01 or .005 to account for convincing strength in the statistics. Furthermore, it is a good idea that one also uses some alternative non-parametric test to see if the results match each other to develop a strong case for the solution. Develop a measurement perspective that will yield value-added and appealing research results. Even if the LS does not meet the criteria given above and the researcher still insists on using FA without compromising statistical robustness, Rasch Analysis is a technique that can be used to convert Likert Scale into true interval items (Sick, 2009; Weaver, 2010). Researcher should be encouraged to used use non-parametric alternatives for FA like CATPCA if the FA solutions are not very good. This is far more preferable than insisting on the use of parametric statistical techniques without establishing their validity. Unlike common belief, it should be noted that non-parametric tests are not at all inferior to parametric tests (Knapp, 1990) and when population is not normal, they even tend to be superior to parametric tests.

Conclusion

In social sciences LS is one of the most popular choices in collecting data on opinion and behavior. On the other hand, FA is frequently used on LS to test the validity of constructs through data reduction or even collapsing cases with high multicollinearity. However, the widespread use of FA on LS has resulted in debates about the legality of FA on LS primarily because LS is debated to be a non-continuous scale and FA is a parametric technique. This paper gives some technical suggestions as to how FA on LS can better be used to improve the statistical appropriateness.

At first, we established that FA can be used on LS considering the utility it generates in social sciences but at the same time look for a somewhat more unbiased solution. The suggestion include: (1) increasing the response categories to 7 or 9 rather than 5, (2) using a p-value of 0.01 or .005 instead of 0.05 that is the default option in many software, (3) using Rasch analysis to transform Likert scale into true interval scale, (4) increasing sample size to at least 300 cases in order to improve the normality of data, provides for low values of loadings and communalities, (5) using polychoric correlation rather than default

Pearson correlation, (6) employing special-purpose software like FACTOR that uses polychoric correlation instead of Pearson correlation, (7) instead of using eigenvalue criterion of greater than 1, the use of scree plot with some subjectivity supported by theory being more appropriate, (8) oblique rotation methods than orthogonal rotations, (9) in case of bad FA solution, resorting to non-parametric options like CATPCA and (10) validating the FA solution with some alternative non-parametric techniques.

There are certain limitations of this paper. At first, it is only a review of literature on FA and LS, it lacks empirical support. Secondly, the paper only tries to explore options for a better FA solution on LS, it does not end the debate on the subject as to the appropriateness of the technique itself. The paper is meant for a functional use and allows researchers to get most out of FA on LS. One might argue that in the presence of several good alternative techniques why use FA. We reckon that the point is worth considering and here we just offer our suggestions only if a researcher insists on employing FA on LS. Finally, this review is not a conclusive work, and leaves room for questions more specifically about sample size, factor loadings and communalities.

References

Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality Progress*, *40*(7), 64–65.

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., & Cochran, J. J. (2020). *Essentials of modern business statistics with microsoft excel*. Cengage Learning.

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Educational Testing Service.

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, *24*(4), 468-491.

Bacchetti, P., Deeks, S. G., & McCune, J. M. (2011). Breaking free of sample size dogma to perform innovative translational research. *Science Translational Medicine*, *3*(87). doi: 10.1126/scitranslmed.3001628

Balasubramanian, N. (2012). Likert technique of attitude scale construction in nursing research. *Asian Journal of Nursing Education and Research*, *2*(2), 65-69.

Barua, A., et al. (2013). Methods for decision-making in survey questionnaires based on likert scale. *Journal of Asian Scientific Research*, *3*(1), 35–38.

Beavers, A. S., Lounsbury, J. W., Richards, J. K., & Huck, S. W. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, *18*(6), 1-13.

Bishop, P. A., & Herron, R. L. (2015). Use and misuse of the likert item responses and other ordinal measures. *International Journal of Exercise Science*, *8*(3), 297-302.

Boone, H. N., & Boone, D. A. (2012). Analyzing likert data. *Journal of Extension*, *50*(2), 1–5.

Brown, J. D. (2011). Likert items and scales of measurement. *Statistics*, *15*(1), 10–14.

Byrne, B. M. (2006). Structural equation modeling with eqs: Basic concepts. *Applications, And Programming: Lawrence Erlbaum Associates*.

Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *Journal of Social Sciences*, *3*(3), 106–116.

Clason, D. L., & Dormody, T. J. (1994). Analyzing data measured by individual likert-type items. *Journal of Agricultural Education*, *35*(4), 31-35.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(1), 1-9.

Doane, D. P., & Seward, L. E. (2011). Measuring skewness: a forgotten statistic? *Journal of Statistics Education*, *19*(2), 1-18.

Dolnicar, S., Grün, B., & Leisch, F. (2016). Increasing sample size compensates for data problems in segmentation studies. *Journal of Business Research*, *69*(2), 992–999.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G. (2014). *The basics of financial econometrics: Tools, concepts, and asset management applications*. John Wiley & Sons.

Garson, G. (2010). Structural equation modeling.[online] url: http://faculty. chass. ncsu. edu/garson. *PT765/structure. htm.[March, 22]*.

Goretzko, D., Pham, T. T. H., & Bühner, M. (2019). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, 1–12.

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, *25*(1), 33–39.

Gorsuch, R. L. (2003). Factor analysis. weiner, i. *Handbook of Psychology: Research Methods in Psychology*, 143–164.

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.

Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, *65*(2), 202–226.

Jakobsson, U. (2004). Statistical presentation and analysis of ordinal data in nursing research. *Scandinavian Journal of Caring Sciences*, *18*(4), 437–440.

Jamieson, S. (2004). Likert scales: How to (ab) use them? *Medical Education*, *38*(12), 1217–1218.

KILIÇ, A. F. (n.d.). Exploratory factor analysis with r software. *Anadolu Üniversitesi Eğitim Fakültesi Dergisi*, *4*(3), 276–293.

Kline, P. (1979). *Psychometrics and psychology*. London: Academic Press.

Knapp, T. R. (1990). Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing research*, *39*(2), 121–123.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as statistical method* (Tech. Rep.).

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140).

Lubke, G. H., & Muthén, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to likert scale data complicates meaningful group comparisons. *Structural Equation Modeling*, *11*(4), 514–534.

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, *23*(1), 151–169.

Marascuilo, L. A., & Levin, J. R. (1983). *Multivariate statistics in the social sciences: A researcher's guide*. Wadsworth Publishing Company.

Mulaik, S. A. (2009). *Foundations of factor analysis*. CRC press.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks: Sage Publications, Inc.

Schreiber, J. B. (2020). Issues and recommendations for exploratory factor analysis and principal component analysis. *Research in Social and Administrative Pharmacy*. doi: 10.1016/j.sapharm.2020.07.027

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, *31*(12), 1428-1441.

Sick, J. (2009). Rasch measurement in language education part 3: The family of rasch models. *Shiken: JALT Testing & Evaluation SIG Newsletter*, *13*(1), 4–10.

Suhr, D. D. (2006). *Exploratory or confirmatory factor analysis?* Cary: SAS Institute, Inc.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics.* Needham Heights: Allyn & Bacon.

Thomas, H. (1982). IQ, interval scales, and normal distributions. *Psychological Bulletin*, *91*(1), 198-202.

Utts, J. M. (2014). *Seeing through statistics*. Cengage Learning.

Vigderhous, G. (1977). The level of measurement and "permissible" statistical analysis in social research. *Pacific Sociological Review*, *20*(1), 61–72.

Vogt, W. P. (1999). *Dictionary for statistics & methodology: A ontechnical guide for the social sciences sage publications*. Thousand Oaks, CA.

Warmbrod, J. R. (2014). Reporting and interpreting scores derived from likert-type scales. *Journal of Agricultural Education*, *55*(5), 30–47.

Weaver, C. T. (2010). *Japanese university students'willingness to use english with different interlocutors* (Unpublished doctoral dissertation). Temple University Libraries.

Wheelwright, M., Levine, Z., Garfinkel-Castro, A., Bushman, T., & Brewer, S. C. (2020). Principal component and factor analysis. In *Advanced quantitative research methods for urban planners.* Routledge.

Zöller, M. (2012). A comparison between principal component analysis and factor analysis.. Retrieved from `http://www.ieee-student-conference.de/fileadmin/papers/2012/ieeegsc2012_submission_1.pdf`