ISSN: 1309 – 6575
**Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi**
Journal of Measurement and Evaluation in Education and Psychology
Technical Brief; 2024; 15(4); 446-451

# FAfA: Factor Analysis for All An R Package to Conduct Factor Analysis with R Shiny Application

Abdullah Faruk KILIÇ *

**Abstract**

This paper presents a presentation of FAfA (the R Shiny application), which was specifically developed for performing complete factor analysis processes. These procedures include data wrangling, assumption checks for exploratory and confirmatory factor analysis, reliability analysis, exploratory graphic analysis, and item weighting. The objective of the paper is to provide users with clear instructions on how to effectively use the FAfA package, therefore guaranteeing precise and consistent outcomes in their research. The FAfA application's primary goal is to integrate EFA and CFA into a single software. Furthermore, FAfA possesses the capability to compute several reliability coefficients related to internal consistency. It can also be utilized when item weighing is desired. This package is advantageous as it enables the verification of assumptions prior to analysis within a single program, facilitates both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) in one application, provides reliability coefficients not accessible in user-interface programs (such as stratified alpha), and integrates exploratory graph analysis, which has rapidly advanced in recent years, into a unified application.

*Keywords: factor analysis, reliability analysis, item weighting, exploratory graph analysis*

## Introduction

Factor analysis (FA) is a widely used method to collect validity evidence for measures. There are numerous software tools available for conducting factor analysis (FA). SPSS can conduct exploratory factor analysis (EFA), Mplus and AMOS can conduct confirmatory factor analysis (CFA), and Factor software is used for EFA. However, there are limitations to the software. First, no module in the software allows for a stand-alone examination of EFA or CFA assumptions. Furthermore, it is typically unfeasible to conduct both EFA and CFA using a single software. While JASP or JAMOVI provide a solution to this challenge, it is essential to note that verifying the assumptions in these software platforms also requires a significant amount of time. In contrast, the FAfA R Shiny application is designed to save researchers time, allowing them to focus on their analysis and interpretation. Furthermore, when conducting EFA in SPSS, the Pearson correlation matrix is utilized. However, due to the increased collection of ordinal data in domains like education and psychology, it may be necessary to use a polychoric correlation matrix. In addition, removing outliers in this software (JAMOVI and JASP) is complicated. So, I created an RShiny app named FAfA to conduct EFA, CFA, and reliability analysis with data wrangling and assumption check properties.

### Dependencies of FAfA Application

In any Shiny application, it is essential to ensure that all necessary packages are loaded. FAfA uses shiny (Chang et al., 2024) for building the user interface, *dplyr* (Wickham et al., 2023) for data manipulation, *psych* for EFA (Revelle, 2024), *lavaan* (Rosseel, 2012) for CFA, *EGAnet* (Golino & Alexander, 2023) for exploratory graph analysis. I used the *psych* (Revelle, 2024) for the alpha coefficient, *MBESS* (Kelley, 2023) for omega, *semTools* (Jorgensen et al., 2022) for structure reliability, *sirt* (Robitzsch, 2021) for stratified alpha, and I wrote code for Armor's theta reliability coefficient. Loading these packages at the beginning ensures that all dependencies are available when needed, which is a best practice in R programming.

* Associate Professor, Trakya University, Faculty of Education, Edirne-Türkiye, afarukkilic@trakya.edu.tr, ORCID ID: 0000-0003-3129-1763

To cite this article:
Kılıç, A.F. (2024). FAfA: Factor analysis for all an R package to conduct factor analysis with R Shiny application*. Journal of Measurement and Evaluation in Education and Psychology*, *15*(4), 446-451. https://doi.org/10.21031/epod.1555805
*Received: 25.09.2024*
*Accepted: 17.12.2024*

### Reading User-Uploaded Data

FAfA analyzes the data formatted by ".dat," ".txt" (Text Document—MS-DOS Format), or ".prn" file. Variable names should not be included in the data set, and missing data should be indicated with NA. The application ensures that the uploaded data is read correctly and is ready for subsequent processing steps. Figure 1 illustrates the data example and the application's control result.

**Figure 1**

*Data Set and Control Results of Data*

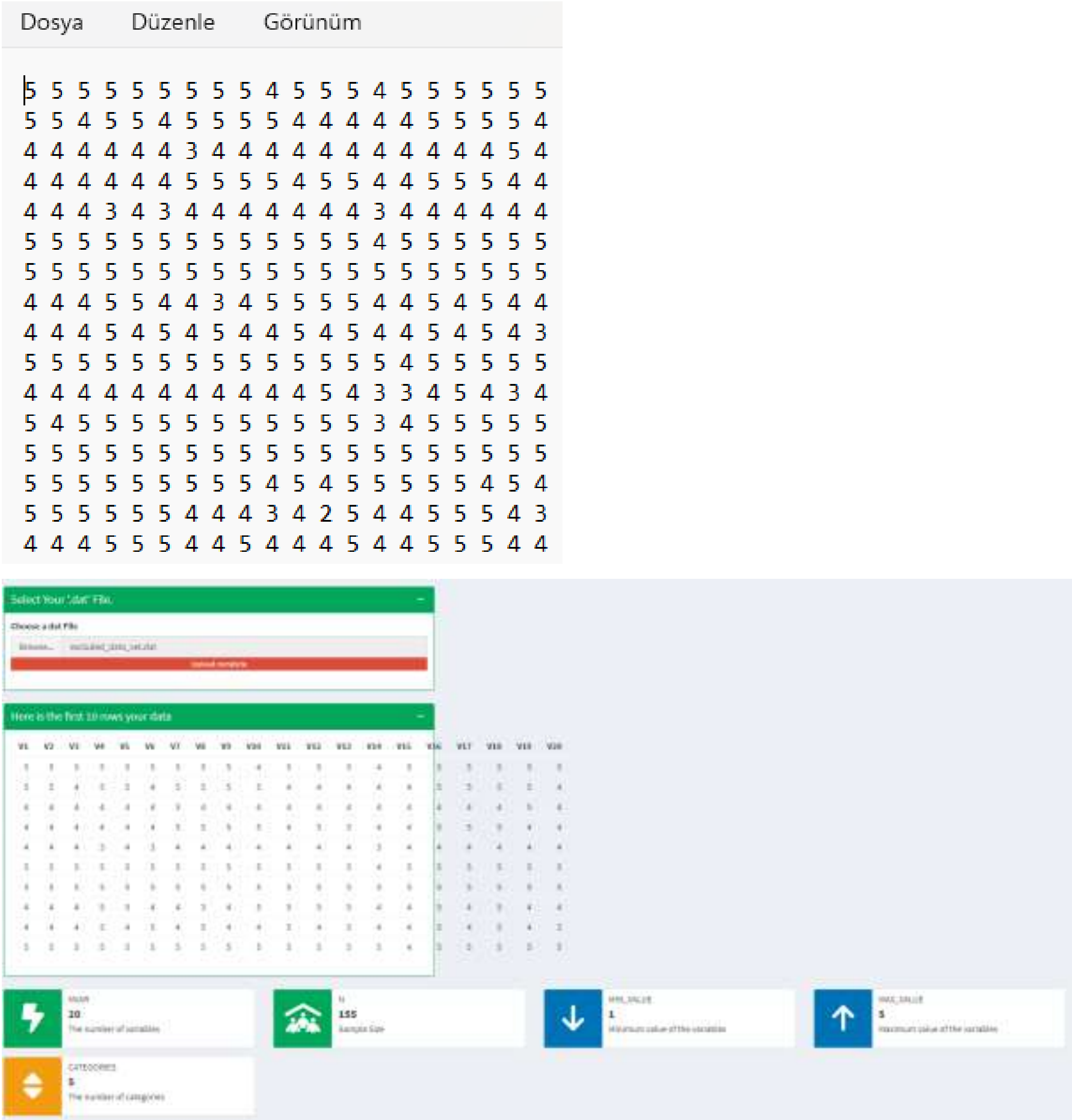


Figure 1 displays the view of the data set and the results of FAfA's check of the data set, which was performed after reading the data in the FAfA application. The dataset has a ".dat" extension, and there are no variable names (column names). After reading the dataset, the first 10 rows are identified as FAfA output. Then, the number of variables in the dataset, sample size, minimum value and maximum value in the dataset, and number of categories are displayed. Thus, it can be examined whether the data set is read correctly or not.

## Wrangling Data

In order to exclude the variables, the column numbers of the variables to be excluded should be written by placing a comma between them. Once you have defined the variables in this manner, use the "Exclude the variables button" to exclude them. Next, save the data set containing the excluded variables to your computer using the download excluded data button, and then read it again in the FAfA application. FAfA can randomly split the data set into two. To accomplish this, first split the data set into two using the "split my data" button, then "download the EFA data" to save the first half on your computer, and download the CFA data to save the second half. Similarly, whatever data set is analyzed should be read again in the FAfA application. This should not be interpreted as applicable to every data set. In scale development research, data is collected again following the examination of the scale's structure using EFA, after which Confirmatory Factor Analysis (CFA) is conducted. Utilize this part to circumvent conducting EFA and CFA on a singular data set, as indicated by Fokkema and Greiff (2017). FAfA uses the Mahalanobis distance statistic to identify multivariate outliers. Accordingly, those that are significant at the a=0.001 level are examined in the Examined Outliers section. This section reports how many outliers there are in the data set. Additionally, you can remove outliers from the data set by using the "Remove outliers from my data" button. Next, save the data set without outliers to your computer using the "Download the data set without outliers" button, and then read it back into FAfA.

## Excluding Variables

The FAfA application incorporates the capability to eliminate user-specified variables from the dataset. Users can designate specific variables to exclude. Excluding variables is often essential for a variety of reasons, including eliminating irrelevant variables, correcting multicollinearity, managing missing data, or reducing item dropouts, which are consequences of EFA. This function also aids in data preparation for more sophisticated studies by eliminating potential sources of bias or noise.

The method of removing variables entails enabling users to input indices (column numbers) to be eliminated from the dataset. Subsequently, the application generates a new dataset by removing the given variables. The new dataset must be uploaded to the application again.

## Assumptions

In the provided code, the assumption function performs several statistical tests and calculations to check the assumptions required for EFA and CFA. This function is a crucial component of the application, as it ensures that the data meets the necessary criteria for valid statistical analysis. The function finds outliers, checks for multicollinearity, and checks for multivariate normality using Mardia's multivariate skewness and kurtosis values (Mardia, 1970). It also finds out what the minimum and maximum values are and the number of missing data points.

The application generates descriptive statistics and checks assumptions for further analysis. This includes calculating various descriptive measures, checking for multicollinearity, and assessing normality. The results are displayed in the application, and the user can download them. Providing descriptive statistics and assumption checks helps users gain a comprehensive understanding of their data. It allows them to identify potential issues and make informed decisions about the appropriate analysis techniques to use. By incorporating these functionalities into the FAfA, users can perform thorough and reliable data analyses.

## Exploratory Graph Analysis (EGA)

Exploratory Graph Analysis (EGA) is a technique used to identify the data's underlying structure by estimating the network structure. EGA is a novel technique introduced in the field of network psychometrics to determine the number of factors that underlie multivariate data. EGA generates a network plot that provides a visual representation of the optimal number of dimensions to keep. Additionally, this plot reveals the clustering of items and the strength of their associations (Golino et al., 2020). The FAfA application performs EGA and provides a visualization of the network, helping users identify clusters or groups of variables that are correlated. Figure 2 demonstrates the EGA results of an example data set.

**Figure 2**

*EGA Results of the FAfA Package*

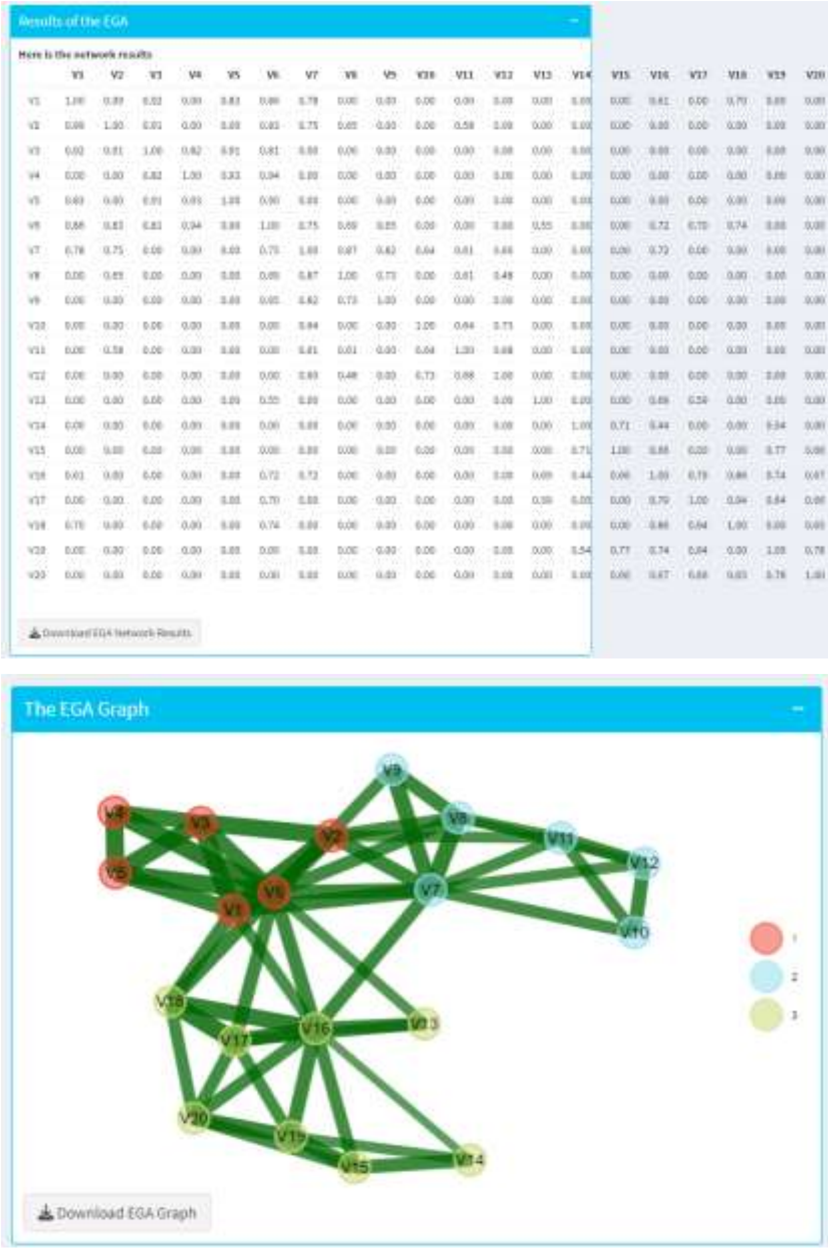


Figure 2 illustrates that FAfA initially presents the outcomes derived from the EGA analysis, followed by the corresponding network graph. The example demonstrates a three-factor structure and strong correlations among the items.

## Exploratory Factor Analysis (EFA)

Exploratory Factor Analysis (EFA) is a statistical technique used to identify the underlying factor structure of the data (Gorsuch, 1974). The FAfA application is utilized for EFA and employs multiple techniques, such as parallel analysis (Horn, 1965) and the Hull approach (Lorenzo-Seva et al., 2011), to ascertain the number of factors. Furthermore, the KMO statistic and the findings of Bartlett's test of sphericity, which are commonly assessed for EFA. FAfA also computes the MSA index, as proposed by Lorenzo-Seva and Ferrando (2021), for each item. Subsequently, FAfA will depict the correlation matrix among the items utilizing a heat map. The results of the factor analysis, including factor loadings and explained variance, are displayed and can be downloaded by the user. Users can conduct the EFA with Pearson or a polychoric/tetrachoric correlation matrix. Various extraction methods, such as principal axis factoring or maximum likelihood estimation, can be used. Oblique and orthogonal rotation methods, such as varimax or oblimin, help achieve a more straightforward factor structure by maximizing the variance explained by each factor. The FAfA provides options for users to choose the extraction and rotation methods, ensuring flexibility in the analysis.

### Confirmatory Factor Analysis (CFA)

Confirmatory Factor Analysis (CFA) validates the factor structure known prior to the analysis (Brown, 2015. Users can define their factor structure, and the FAfA estimates the model and provides various fit measures to assess the model fit. The results, including factor loadings and modification indices, are displayed in the FAfA output. Users can download these outputs. The CFA process entails specifying a factor model in which the relationships between observed variables and latent factors are defined. The model is then estimated using techniques such as maximum likelihood estimation. FAfA reports fit measures such as Chi Square, degrees of freedom (df), Chi Square/df, p value of Chi Square, CFI, TLI, RFI, SRMR, RMSEA and its 90% confidence interval. They are used to assess the model's fit to the data. The application provides a summary of these fit measures, helping users evaluate the adequacy of their specified factor model. In addition, model modification suggestions can be examined.

### Reliability Analysis

Reliability analysis assesses the internal consistency of the data, providing a measure of the reliability of the scales used (Mueller & Knapp, 2019). The FAfA calculates various reliability coefficients, such as Cronbach's alpha and McDonald's omega. The results are displayed, and the user can download them. Assessing reliability is essential in psychometrics. FAfA calculates not only the unidimensional reliability coefficients, but also a multidimensional reliability coefficient named stratified alpha. Stratified alpha has been proposed to calculate the reliability of composite scores obtained from measurement tools with multiple subdimensions (Cronbach et al., 1965).

### Item Weighting

I added the item weighting function to enhance the validity of the measures suggested by Kılıç and Doğan (2019). Item weighting is a valuable technique for refining measurement scales. By adjusting scores based on item properties, users can enhance the construct validity and reliability of their scales. The FAfA provides a convenient tool for implementing item weighting, helping users improve their measurement instruments. This weighting function assigns a weight to the item based on the combined values of item difficulty and the respondent's average score. If this sum exceeds 1, the item reliability is incorporated into the respondent's answers. If this sum does not exceed 1, the respondent's score remains unchanged (1 for a correct item, 0 for an incorrect item).

## Conclusion

This manuscript provides a detailed review of an R Shiny application named FAfA. This comprehensive review serves as a guide for researchers and practitioners looking to utilize FAfA for their data analysis, enhancing the accessibility and usability of advanced statistical techniques. The R package FAfA is available on the Comprehensive R Archive Network (CRAN; http://www.cran.r-project.org). FAfA can be installed with `install.packages("FAfA", dependency = TRUE)` code in R. The FAfA package requires additional packages. The FAfA package specifies these packages in its suggestions section. If you get an error after running that installation code, running the following code may also help you prevent errors.

```
install.packages(c("EFA.MRFA", "EFA.dimensions", "EFAtools", "EGAnet", "MB
ESS", "config", "dplyr", "energy", "ggcorrplot", "golem", "lavaan", "mctes
t", "moments", "mvnormalTest", "pastecs", "psych", "psychometric", "semPlo
t", "semTools", "shiny", "shinycssloaders", "shinydashboard", "sirt"))
```

After loading the packages with these scripts, we may invoke the package using the `install.packages(FafA)`. A function exists within the package. This function is the `run_app()` function. The `run_app()` function executes the package directly in English. The function `run_app(lang = "tr")` can be utilized to activate Turkish language support if preferred. Source code and documentation are freely available from https://CRAN.R-project.org/package=FAfA.

## Declarations

**Conflict of Interest:** No potential conflict of interest was reported by the author.

**Ethical Approval:** I confirm that I have followed all ethical guidelines for authorship. This study does not necessitate ethical approval due to its nature as a software presentation.

## References


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

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2024). *shiny: Web Application Framework for R.* (Version 1.7.5) [R package]. https://CRAN.R-project.org/package=shiny

Cronbach, L. J., Schönemann, P., & McKie, D. (1965). Alpha coefficients for stratified-parallel tests. *Educational and Psychological Measurement, 25*(2), 291–312. https://doi.org/10.1177/001316446502500201

Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble. *European Journal of Psychological Assessment, 33*(6), 399–402. https://doi.org/10.1027/1015-5759/a000460

Golino, H., & Alexander, C. P. (2023). *EGAnet: Exploratory Graph Analysis—A framework for estimating the number of dimension in multivariate data using network psychometrics.* (Version 2.1.0) [R package]. https://cran.r-project.org/web/packages/EGAnet/EGAnet.pdf

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods, 25*(3), 292–320. https://doi.org/10.1037/met0000255

Gorsuch, R. L. (1974). *Factor analysis.* W. B. Saunders.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling* (Version 0.5-6) [R package]. https://CRAN.R-project.org/package=semTools

Kelley, K. (2023). *MBESS.* (Version 4.9.3) [R package]. https://cran.r-project.org/package=MBESS

Kılıç, A. F., & Doğan, N. (2019). The Effect of item weighting on reliability and validity. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 10*(2), 148–163. https://doi.org/10.21031/epod.516057

Lorenzo-Seva, U., & Ferrando, P. J. (2021). MSA: The forgotten index for identifying inappropriate items before computing exploratory item factor analysis. *Methodology, 17*(4), 296–306. https://doi.org/10.5964/meth.7185

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research, 46*(2), 340–364. https://doi.org/10.1080/00273171.2011.564527

Mueller, R. O., & Knapp, T. R. (2019). Reliability and validity. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (2nd. ed., pp. 397–401). Routledge.

Revelle, W. (2024). *psych: Procedures for psychological, psychometric, and personality research* (Version 2.4.3) [R package]. https://cran.r-project.org/package=psych

Robitzsch, A. (2021). *sirt: Supplementary Item Response Theory Models.* (Version 3.9-3) [R package]. https://cran.r-project.org/package=sirt

Rosseel, Y. (2012). lavaan: An r package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Wickham, H. (2021). *Mastering shiny*. O'Reilly Media.

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of data manipulation.* (Version 1.1.3) [R package]. https://CRAN.R-project.org/package=dplyr