

EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools

Markus D. Steiner¹ and Silvia Grieder²

¹ Center for Cognitive and Decision Sciences, Department of Psychology, University of Basel, Switzerland ² Division of Developmental and Personality Psychology, Department of Psychology, University of Basel, Switzerland

DOI: [10.21105/joss.02521](https://doi.org/10.21105/joss.02521)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Frederick Boehm](#) ↗

Reviewers:

- [@jacobsoj](#)
- [@chainsawriot](#)

Submitted: 19 July 2020

Published: 03 August 2020

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

In the social sciences, factor analysis is a widely used tool to identify latent constructs underlying task performance or the answers to questionnaire items. Exploratory factor analysis (EFA) is a data-driven approach to factor analysis and is used to extract a smaller number of common factors that represent or explain the common variance of a larger set of manifest variables (see Costello & Osborne, 2005; and Watkins, 2018 for an overview). Several decisions have to be made in advance when performing an EFA, including the number of factors to extract, and the extraction and rotation method to be used. After a factor solution has been found, especially for data structures in the field of intelligence research where usually high, positive factor intercorrelations occur, it is useful to subject the resulting factor solution to an orthogonalization procedure to achieve a hierarchical factor solution with one general and several specific factors. From this orthogonalized, hierarchical solution, the variance can then be partitioned to estimate the relative importance of the general versus the specific factors using omega reliability coefficients (e.g., McDonald, 1999).

EFAtools is an R package (R Core Team, 2019) that enables fast and flexible analyses in an EFA framework, from tests for suitability of the data for factor analysis and factor retention criteria to hierarchical factor analysis with Schmid-Leiman transformation (Schmid & Leiman, 1957) and McDonald's omegas (e.g., McDonald, 1999). The package's core functionalities are listed in Table 1.

Statement of Need

Compared to other R packages with which EFA can be performed, *EFAtools* has several advantages, including fast implementations using *Rcpp* (Eddelbuettel & Balamuta, 2017; Eddelbuettel & Sanderson, 2014), more flexibility in the adjustment of implementation features, the ability to reproduce the R *psych* (Revelle, 2019) and SPSS (IBM, 2015) implementations of some analyses methods (see vignette *Replicate SPSS and R psych results with EFAtools*), as well as the inclusion of recommended implementations for these methods based on simulation analyses by the authors (pending publication). Finally, the package includes the implementation of the, as of yet, most comprehensive set of factor retention criteria in R, including recently developed criteria such as the Hull method (Lorenzo-Seva, Timmerman, & Kiers, 2011), comparison data (Ruscio & Roche, 2012), and the empirical Kaiser criterion (Braeken & van Assen, 2016). As recommended by Auerwald & Moshagen (2019), multiple factor retention criteria should be examined simultaneously to check their convergence, which now is easily possible with a comprehensive function in *EFAtools* incorporating all implemented factor retention criteria for simultaneous application. Minor advantages over and above the

existing implementations in R include that when intending to perform a Schmid-Leiman transformation, this can be done on an obliquely rotated solution obtained with functions from the *EFAtools* or the *psych* package instead of being forced to perform the whole EFA procedure again. Moreover, our implementation of McDonald's omegas calculations include the possibility of manual variable-to-factor correspondences (as are needed for variance partitioning for predetermined / theoretical composites) in addition to automatically determined variable-to-factor correspondences (as done, for example, in the *psych* package). Further, the *EFAtools* function to compute McDonald's omegas can easily be applied on *EFAtools* and *psych* Schmid-Leiman solutions as well as on *lavaan* (Rosseel, 2012) second-order, bifactor, and single factor solutions (including solutions from multiple group analyses).

Development and Purpose

EFAtools was designed for use in the social sciences in general and is especially suitable for research on cognitive abilities or other hierarchically organized constructs as well as for more time-consuming applications such as in simulation analyses. Its development arose from the need for a tool for easy replication and comparison of EFA solutions from different programs, namely R and SPSS (pending publication by the authors), and has already been used in another publication (Grieder & Grob, 2019). The package was then expanded for a broader, easy, fast, and flexible use of EFA tools such that it is now suitable for most projects within the EFA framework.

Table 1: Core functionalities of *EFAtools*.

| Topic | Method | Function |
|---------------------------------|--|-----------------|
| Suitability for factor analysis | Bartlett's test of sphericity | BARTLETT() |
| | Kaiser-Meyer-Olkin criterion | KMO() |
| Factor retention criteria | Comparison data | CD() |
| | Empirical Kaiser criterion | EKC() |
| | Hull method | HULL() |
| | Kaiser-Guttman criterion | KGC() |
| | Parallel analysis | PARALLEL() |
| | Scree plot | SCREE() |
| | Sequential model tests | SMT() |
| | RMSEA lower bound criterion | SMT() |
| | AIC criterion | SMT() |
| | Factor extraction methods | |
| | Principal axis factoring | EFA() |
| | Maximum likelihood | EFA() |
| | Unweighted least squares | EFA() |
| Rotation methods | Orthogonal: Varimax, equamax, quartimax, geominT, bentlerT, bifactorT | EFA() |
| | Oblique: Promax, oblimin, quartimin, simplimax, bentlerQ, geominQ, bifactorQ | EFA() |
| Factor scores | Different methods for calculating factor scores | FACTOR_SCORES() |
| Hierarchical factor analysis | Schmid-Leiman transformation | SL() |
| | McDonald's omegas | OMEGA() |

Note. All functions for suitability for factor analysis and factor retention criteria can be called in any desired combination using the `N_FACTORS()` function.

Installation

The *EFAtools* package can be installed from CRAN using `install.packages("EFAtools")`. Moreover, the development version can be installed from GitHub (<https://github.com/mdsteiner/EFAtools>) using `devtools::install_github("mdsteiner/EFAtools", build_vignettes = TRUE)`.

Acknowledgements

We thank Dirk Wulff for helpful suggestions concerning the C++ implementations.

References

- 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. doi:[10.1037/met0000200](https://doi.org/10.1037/met0000200)
- Braeken, J., & van Assen, M. A. L. M. (2016). An empirical Kaiser criterion. *Psychological Methods*, 22(3), 450–466. doi:[10.1037/met0000074](https://doi.org/10.1037/met0000074)
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10, 1–9. doi:[10.7275/jyj1-4868](https://doi.org/10.7275/jyj1-4868)
- Eddelbuettel, D., & Balamuta, J. J. (2017). Extending R with C++: A brief introduction to Rcpp. *PeerJ Preprints*, 5, e3188v1. doi:[10.7287/peerj.preprints.3188v1](https://doi.org/10.7287/peerj.preprints.3188v1)
- Eddelbuettel, D., & Sanderson, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71, 1054–1063. doi:[10.1016/j.csda.2013.02.005](https://doi.org/10.1016/j.csda.2013.02.005)
- Grieder, S., & Grob, A. (2019). Exploratory factor analysis of the Intelligence and Development Scales–2: Implications for theory and practice. *Assessment*. Advance online publication. doi:[10.1177/1073191119845051](https://doi.org/10.1177/1073191119845051)
- IBM, C. (2015). *IBM SPSS Statistics for Macintosh, Version 23.0*. Armonk, NY.
- 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. doi:[10.1080/00273171.2011.564527](https://doi.org/10.1080/00273171.2011.564527)
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum. ISBN: [978-1-4106-0108-7](https://www.isbn-international.org/product/978-1-4106-0108-7)
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Revelle, W. (2019). *Psych: Procedures for psychological, psychometric, and personality research*. Evanston, IL: Northwestern University. Retrieved from <https://CRAN.R-project.org/package=psych>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. doi:[10.18637/jss.v048.i02](https://doi.org/10.18637/jss.v048.i02)
- Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, 24(2), 282–292. doi:[10.1037/a0025697](https://doi.org/10.1037/a0025697)

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53–61. doi:[10.1007/BF02289209](https://doi.org/10.1007/BF02289209)

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44, 219–246. doi:[10.1177/0095798418771807](https://doi.org/10.1177/0095798418771807)