

Editorial of the *Psych* Special Issue “Computational Aspects, Statistical Algorithms and Software in Psychometrics”

Alexander Robitzsch ^{1,2} 

¹ IPN—Leibniz Institute for Science and Mathematics Education, Olshausenstraße 62, 24118 Kiel, Germany; robitzsch@leibniz-ipn.de

² Centre for International Student Assessment (ZIB), Olshausenstraße 62, 24118 Kiel, Germany

1. Introduction

Statistical software in psychometrics has made tremendous progress in providing open source solutions (e.g., software R, Julia, Python). In the articles of this Special Issue, focus was devoted to computational aspects and statistical algorithms for psychometric methods: for example, shared experiences about efficient implementation aspects or how to handle vast datasets in psychometric modeling were discussed in detail. On the other hand, articles that introduce new software packages were also published. Furthermore, there were several software tutorials that could prove helpful for applied practitioners. The discussed psychometric models included structural equation models, multilevel models, item response models, cognitive diagnostic models, missing data models, and machine learning methods.

I would like to thank all authors of the 31 articles of this Special Issue for their excellent contributions that provided a perfect fit to the scope of the Special Issue. Moreover, I would like to sincerely thank all reviewers, handling editors, and the editorial staff of *Psych* for their support.

The rest of this editorial gives a brief overview of the published articles.

2. Articles in This Special Issue

In the following, I classified the articles according to five categories. Each category is treated in a subsection.

2.1. Multilevel Modeling and Structural Equation Modeling

The article of Rosseel [1] discusses maximum likelihood estimation for two-level structural equation models under a perspective of computationally efficient implementations of the observed log-likelihood function. By presenting R snippets, several implementations are compared that motivate the final implementation in the *lavaan* package.

Jak et al. [2] discuss the estimation of different two-level factor models for cluster-level constructs in the software package *lavaan* and *Mplus*. They compare the so-called configural model and the simultaneous shared-and-configural model to replicate the simulation study of Stapleton and Johnson (2019, *J. Educ. Behav. Stat.*). As an outcome of their study, Jak et al. [2] worried about default settings in the *Mplus* software for the chi-square test of model fit and provide suggestions for circumventing these issues.

As a comment on Jak et al. [2], the *Mplus* authors Asparouhov and Muthén [3] suggest modifying of the robust chi-square test of fit. The improved statistic yielded more accurate type I error rates when the estimated model parameters are at the boundary of the admissible parameter space, which was the focus in Jak et al. [2].

Hecht et al. [4] investigate different Markov-chain Monte Carlo implementations of the two-level random intercept model in the popular general Bayesian software packages JAGS and Stan. The authors compare a parameterization based on sufficient statistics (i.e., means and covariances; covariance- and mean-based parametrization) with a classic parameterization that also samples random effects. Computational efficiency was assessed as the effective



Citation: Robitzsch, A. Editorial of the *Psych* Special Issue

“Computational Aspects, Statistical Algorithms and Software in Psychometrics”. *Psych* **2022**, *4*, 114–118. <https://doi.org/10.3390/psych4010011>

Received: 21 February 2022

Accepted: 25 February 2022

Published: 2 March 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

sample size per second. It turned out that Stan outperformed JAGS in the covariance- and mean-based parameterizations, but JAGS outperformed Stan in the classic parameterization.

Zitzmann et al. [5] discuss the assessment of the convergence of Markov-chain Monte Carlo estimation in the *Mplus* software. They argue that the effective sample size should be preferred over the frequently used potential scale reduction factor. Zitzmann and Hecht (2019, *Struct. Equ. Modeling*) propose a method that can be used to check whether a minimum effective sample size has been reached in *Mplus*. This method was evaluated in a simulation study in the contribution of this Special Issue.

Schoemann and Jorgensen [6] review methods of estimating and testing latent variable interactions in structural equation modeling, with a focus on the product indicator method. They demonstrate how the product indicator methods of examining latent interactions can provide an accurate method to estimate and test latent interactions. Moreover, the authors show how this method can be implemented in any structural equation modeling software package. Schoemann and Jorgensen [6] illustrate the implementation of the product indicator method in the *semTools* package that relies on the R package *lavaan* for fitting the structural equation model.

Jorgensen [7] show how to use structural equation modeling for estimating error components in generalizability theory for continuous and ordinal items. The author uses real and simulated datasets to demonstrate how a structural equation model can be specified to estimate the absolute error by posing constraints on the mean structure (for continuous items) as well as the thresholds (for ordinal items). Different estimators for continuous and ordinal items are compared using the R packages *lavaan* and *gtheory*.

The article of Arnold et al. [8] investigates parameter heterogeneity with respect to covariates in structural equation models. The authors demonstrate how the individual parameter contribution regression framework could be used to predict differences in any parameter of a structural equation model. Arnold et al. [8] implement the individual parameter regression framework in the R package *ipcr*. Furthermore, they compare the performance of individual parameter regression with alternative methods for dealing with parameter heterogeneity (e.g., regularization methods, structural equation models with interaction effects) in a simulation study.

Li et al. [9] provide a tutorial on the sparse estimation of structural equation models (i.e., regularized structural equation modeling). Regularization techniques penalize the complexity of the model and can perform parameter selection in an automatic and completely data-driven way. Li et al. [9] illustrate regularized structural equation modeling using a detailed example code in the R package *regsem*.

Christensen and Golino [10] investigate the assessment of sampling variability in exploratory graph analysis with a bootstrap approach. They conduct a simulation study to assess the suitability of several sampling statistics (i.e., descriptive statistics, structural consistency estimates, item stability statistics). Moreover, Christensen and Golino [10] illustrate their method in the R package *EGAnet*.

2.2. Item Response Modeling and Categorical Data Modeling

Beisemann et al. [11] compare several acceleration methods for the expectation-maximization (EM) algorithm that is often prone to slow convergence. The acceleration techniques for the EM algorithm were applied to marginal maximum likelihood estimation of item response models and mixture models. Beisemann et al. [11] showed that all three studied acceleration methods reduced the number of total log-likelihood evaluations. Hence, using them might be an important part of the implementation of efficient software.

Garnier-Villarreal et al. [12] compare different estimation methods for multidimensional item response models in a large simulation study. They compare limited information methods such as implemented in *lavaan*, marginal maximum likelihood estimation in *mirt*, and Markov chain Monte Carlo estimation in the Stan software. The study of Garnier-Villarreal et al. [12] provides recommendations for applied researchers on which estimation methods should be preferred in particular data-generating constellations.

Ulitzsch and Nestler [13] also focus on estimating multidimensional item response models. The authors compare Markov-chain Monte Carlo estimation in Stan and marginal maximum likelihood estimation in the `TAM` package with variational Bayes estimation implemented in Stan. Ulitzsch and Nestler [13] conclude that variational Bayes was computationally much more efficient than Markov-chain Monte Carlo estimation but did not outperform marginal maximum likelihood estimation. Moreover, because variational Bayes estimates provide biased estimates of item discriminations, the authors argue that variational Bayes is not a viable alternative for estimating multidimensional item response models.

In the article of Kolbe et al. [14], the association of two ordinal variables by means of polychoric correlations is studied. They show that the estimated polychoric correlation is biased if the underlying continuous latent variable is not bivariate and normally distributed. Kolbe et al. [14] illustrate how various bivariate distributions could be fitted to ordinal data and examined how estimates of the polychoric correlation may vary under different distributional assumptions. As a conclusion, the authors noted that the bivariate normal or the bivariate skew-normal distribution might only rarely hold in empirical datasets.

Bulut et al. [15] is a tutorial paper of the `eirm` package that implements exploratory item response models. The functionality of the `eirm` package includes traditional item response models (e.g., Rasch model, partial credit model, and rating scale model), item-explanatory models (i.e., a linear logistic test model), and person-explanatory models (i.e., latent regression models) for both dichotomous and polytomous responses. Bulut et al. [15] illustrate the general functionality of the `eirm` package with annotated R codes based on the Rosenberg self-esteem scale as a running empirical example.

Finnemann et al. [16] is an introduction to the Ising model. They provided a conceptual introduction with a survey of Ising-related software packages in R. The authors use simulation studies to assess how the Ising model captures local-alignment dynamics. In the article, Finnemann et al. [16] offer recommendations on when to use frequentist or Bayesian estimation for the Ising model.

The article of Feuerstahler [17] is a tutorial paper for the `flexmet` package that estimates the filtered monotonic polynomial item response model for dichotomous and polytomous items. This model is a semiparametric item response model that allows for more flexible function shapes and includes traditional item response models as special cases. The tutorial of Feuerstahler [17] aims at providing both an introduction to the unique features of the filtered polynomial model and a guide to its implementation in the R package `flexmet`.

Debelak and Debeer [18] conduct a simulation study on detecting differential item functioning (DIF) for continuous covariates in multistage tests. The authors implement a linear logistic regression test and two score-based DIF tests in the R package `mstDIF`. It turned out that the score-based tests had larger power against DIF effects than the linear logistic regression test.

Shi et al. [19] show how to perform the analysis of a G-DINA model in the R packages `GDINA`, `CDM`, and `cdmTools`. The G-DINA model framework is central to the literature of cognitive diagnostic modeling. The article provides an overview of several typical steps that are conducted in a G-DINA analysis: Q-matrix evaluation, estimation of the G-DINA model, model fit evaluation, item diagnosticity investigation, estimation of classification reliability, and the presentation and visualization of results.

Sorrel et al. [20] provide an overview of recent developments in cognitive diagnosis computerized adaptive testing implemented in the R package `cacatR`. The package includes functionalities for data generation, model selection based on relative fit information, implementation of several item selection rules such as item exposure control, and the evaluation of performance in terms of classification accuracy, item exposure, and test length.

Heine and Stemmler [21] present the application configural frequency analysis in the R package `confreq`. The configural frequency analysis is a person-centered approach that analyzes the residuals of non-fitting models. The authors presented different kinds of configural frequency analyses: the first-order configural frequency analysis based on the null hypothesis of independence, configural frequency analysis with covariates, and the two-

sample configural frequency analysis. Heine and Stemmler [21] illustrate the estimation with R code using the `confreq` package.

2.3. Missing Data and Synthetic Data

Keller [22] provides a brief overview of the factored regression framework (i.e., sequential modeling) for imputing multiple missing data. The author describes the functional notation used to conceptualize the models and generate multiple imputations using this framework within the *Blimp* software. A mediation model with accompanying code is used as an illustration.

Dai [23] reviews the commonly used methods for dealing with missing item responses in psychometrics and examines their performance in a simulation study. Furthermore, the R package `TestDataImputation` is used in an illustration with an example data set.

Volker and Vink [24] outline a workflow for generating synthetic data with the multiple imputation software `mice`. It was demonstrated in a simulation study that the analysis results obtained on synthetic data yielded unbiased and valid statistical inference. Volker and Vink [24] argue that the ease of use when synthesizing data with `mice`, along with the validity of inferences obtained, demonstrates rich possibilities for data dissemination.

2.4. Large-Scale Assessment Methodology

Mirazchiyski [25] introduce the R package `RALSA` (R analyzer for large-scale assessments) for the analysis of international, educational, large-scale assessment data. The article focuses on the technical aspects of `RALSA`. The use of the `data.table` package for memory efficiency, speed, and efficient computations is illustrated using examples. Mirazchiyski [25] mention the utilization of code reuse practices to achieve consistency, efficiency, and safety in the computations performed by the analysis functions of the `RALSA` package.

Becker et al. [26] introduce the R package `eatATA`, which allows the usage of several mixed-integer programming solvers for automated test assembly. The general functionality and the common workflow of `eatATA` are presented using a minimal example and four more complex use cases.

In Gary et al. [27], it is explained how to model norm scores with the R package `cNORM`. The `cNORM` package is designed to determine norm scores when the latent ability to be measured varies with age or other explanatory variables. Gary et al. [27] briefly introduce the statistical modeling behind the implementation and apply their proposed method using a real dataset from a reading comprehension test.

Andersen and Zehner [28] introduce the `shinyReCoR` Shiny app that utilizes a cluster-based method for automatically coding open-ended text responses. The app guides users through the complete workflow such as text corpus compilation, semantic space building, preprocessing of the text data, and clustering.

Ludwig et al. [29] apply a transformer-based approach to automated essay scoring in the Python software and compared it with the bag of words approach. The authors argue that the transformer-based approach has significant advantages, while a bag of words approach suffers from not taking word order into account and reducing the words to their stem. Furthermore, it is demonstrated how such models could improve the accuracy of human ratings.

2.5. Applications and Research Practice

Hartmann et al. [30] introduce the R package `holland`, which enables the computation of the most important descriptive coefficients based on John L. Holland's theory of vocational choice. The article presents an overview of the package and examines its application for research and practice.

Finally, the article of Peikert et al. [31] demonstrates how the R package `repro` can support researchers in creating fully computationally reproducible research projects. Several applications such as the preregistration of research plans with code (i.e., preregistration as code) were provided.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Rosseel, Y. Evaluating the observed log-likelihood function in two-level structural equation modeling with missing data: From formulas to R code. *Psych* **2021**, *3*, 197–232. [\[CrossRef\]](#)
2. Jak, S.; Jorgensen, T.D.; Rosseel, Y. Evaluating cluster-level factor models with lavaan and Mplus. *Psych* **2021**, *3*, 134–152. [\[CrossRef\]](#)
3. Asparouhov, T.; Muthen, B. Robust chi-square in extreme and boundary conditions: Comments on Jak et al. (2021). *Psych* **2021**, *3*, 542–551. [\[CrossRef\]](#)
4. Hecht, M.; Weirich, S.; Zitzmann, S. Comparing the MCMC efficiency of JAGS and Stan for the multi-level intercept-only model in the covariance- and mean-based and classic parametrization. *Psych* **2021**, *3*, 751–779. [\[CrossRef\]](#)
5. Zitzmann, S.; Weirich, S.; Hecht, M. Using the effective sample size as the stopping criterion in Markov chain Monte Carlo with the Bayes module in Mplus. *Psych* **2021**, *3*, 336–347. [\[CrossRef\]](#)
6. Schoemann, A.M.; Jorgensen, T.D. Testing and interpreting latent variable interactions using the semTools package. *Psych* **2021**, *3*, 322–335. [\[CrossRef\]](#)
7. Jorgensen, T.D. How to estimate absolute-error components in structural equation models of generalizability theory. *Psych* **2021**, *3*, 113–133. [\[CrossRef\]](#)
8. Arnold, M.; Brandmaier, A.M.; Voelke, M.C. Predicting differences in model parameters with individual parameter contribution regression using the R package ipcr. *Psych* **2021**, *3*, 360–385. [\[CrossRef\]](#)
9. Li, X.; Jacobucci, R.; Ammerman, B.A. Tutorial on the use of the regsem package in R. *Psych* **2021**, *3*, 579–592. [\[CrossRef\]](#)
10. Christensen, A.P.; Golino, H. Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A Monte Carlo simulation and tutorial. *Psych* **2021**, *3*, 479–500. [\[CrossRef\]](#)
11. Beisemann, M.; Wartlick, O.; Doebler, P. Comparison of recent acceleration techniques for the EM algorithm in one- and two-parameter logistic IRT models. *Psych* **2020**, *2*, 209–252. [\[CrossRef\]](#)
12. Garnier-Villareal, M.; Merkle, E.C.; Magnus, B.E. Between-item multidimensional IRT: How far can the estimation methods go? *Psych* **2021**, *3*, 404–421. [\[CrossRef\]](#)
13. Ulitzsch, E.; Nestler, S. Evaluating Stan’s variational Bayes algorithm for estimating multidimensional IRT models. *Psych* **2022**, *4*, 73–88. [\[CrossRef\]](#)
14. Kolbe, L.; Oort, F.; Jak, S. Bivariate distributions underlying responses to ordinal variables. *Psych* **2021**, *3*, 562–578. [\[CrossRef\]](#)
15. Bulut, O.; Gorgun, G.; Yildirim-Erbasli, S.N. Estimating explanatory extensions of dichotomous and polytomous Rasch models: The eirm package in R. *Psych* **2021**, *3*, 308–321. [\[CrossRef\]](#)
16. Finnemann, A.; Borsboom, D.; Epskamp, S.; van der Maas, H.L.J. The theoretical and statistical Ising model: A practical guide in R. *Psych* **2021**, *3*, 593–617. [\[CrossRef\]](#)
17. Feuerstahler, L. Flexible item response modeling in R with the flexmet package. *Psych* **2021**, *3*, 447–478. [\[CrossRef\]](#)
18. Debelak, R.; Debeer, D. An evaluation of DIF tests in multistage tests for continuous covariates. *Psych* **2021**, *3*, 618–638. [\[CrossRef\]](#)
19. Shi, Q.; Ma, W.; Robitzsch, A.; Sorrel, M.A.; Man, K. Cognitively diagnostic analysis using the G-DINA model in R. *Psych* **2021**, *3*, 812–835. [\[CrossRef\]](#)
20. Sorrel, M.A.; Najera, P.; Abad, F.J. cdcatr: An R package for cognitive diagnostic computerized adaptive testing. *Psych* **2021**, *3*, 386–403. [\[CrossRef\]](#)
21. Heine, J.H.; Stemmler, M. Analysis of categorical data with the R package confreq. *Psych* **2021**, *3*, 522–541. [\[CrossRef\]](#)
22. Keller, B.T. An introduction to factored regression models with Blimp. *Psych* **2022**, *4*, 10–37. [\[CrossRef\]](#)
23. Dai, S. Handling missing responses in psychometrics: Methods and software. *Psych* **2021**, *3*, 673–693. [\[CrossRef\]](#)
24. Volker, T.B.; Vink, G. Anonymized shareable data: Using mice to create and analyze multiply imputed synthetic datasets. *Psych* **2021**, *3*, 703–716. [\[CrossRef\]](#)
25. Mirazchiyski, P.V. RALSA: Design and implementation. *Psych* **2021**, *3*, 233–248. [\[CrossRef\]](#)
26. Becker, B.; Debeer, D.; Sachse, K.A.; Weirich, S. Automated test assembly in R: The eatATA package. *Psych* **2021**, *3*, 96–112. [\[CrossRef\]](#)
27. Gary, S.; Lenhard, W.; Lenhard, A. Modelling norm scores with the cNORM package in R. *Psych* **2021**, *3*, 501–521. [\[CrossRef\]](#)
28. Andersen, N.; Zehner, F. shinyReCoR: A shiny application for automatically coding text responses using R. *Psych* **2021**, *3*, 422–446. [\[CrossRef\]](#)
29. Ludwig, S.; Mayer, C.; Hansen, C.; Eilers, K.; Brandt, S. Automated essay scoring using transformer models. *Psych* **2021**, *3*, 897–915. [\[CrossRef\]](#)
30. Hartmann, F.G.; Heine, J.H.; Ertl, B. Concepts and coefficients based on John L. Holland’s theory of vocational choice—Examining the R package holland. *Psych* **2021**, *3*, 728–750. [\[CrossRef\]](#)
31. Peikert, A.; van Lissa, C.J.; Brandmaier, A.M. Reproducible research in R: A tutorial on how to do the same thing more than once. *Psych* **2021**, *3*, 836–867. [\[CrossRef\]](#)