psych

Editorial

Editorial for the Special Issue "Computational Aspects and Software in Psychometrics II"

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

There has been tremendous progress in statistical software in the field of psychometrics in providing open-source solutions. The focus of this Special Issue, "Computational Aspects and Software in Psychometrics II" (https://www.mdpi.com/journal/psych/special_issues/Computational_Psychometrics; accessed on 5 September 2023), is on computational aspects and statistical algorithms for psychometric methods. The Special Issue covers software articles, as well as simulation studies and review articles, and is a successful continuation of the previous *Psych* Special Issue, "Computational Aspects, Statistical Algorithms and Software in Psychometrics" [1] (https://www.mdpi.com/journal/psych/special_issues/Computational_Algorithms_Psychometrics; accessed on 5 September 2023).

The articles published in this Special Issue are discussed below in chronological order of publication.

The article by Dai and Svetina Valdivia [2], titled "Dealing with missing responses in cognitive diagnostic modeling", examines missing data handling methods for missing dichotomous item responses when the analysis model of interest is the deterministic inputs, noisy "and" gate (DINA) model. The utilized imputation methods were evaluated in terms of classification accuracy. It transpired that no single missing data handling method was superior to all other methods.

The article by Zitzmann, Walther, Hecht, and Nagengast [3], titled "What is the maximum likelihood estimate when the initial solution to the optimization problem is inadmissible? The case of negatively estimated variances", investigates the issue of negatively estimated variances in structural equation modeling. The authors evaluate the strategy whereby an initially estimated negative variance should be set to zero in a final optimization.

The article by Sen and Yildirim [4], titled "A tutorial on how to conduct meta-analysis with IBM SPSS statistics", is a software tutorial for conducting statistical methods for meta-analysis in SPSS software. Given this article's large number of views (i.e., 13,642 views as of 24 August 2023), meta-analysis methods, and their implementation in SPSS, seem to attract many researchers in the field.

The article by Zitzmann [5], titled "A cautionary note regarding multilevel factor score estimates from lavaan", pointed to the implementation issue of factor scores in multilevel models in the popular R package lavaan. It is hoped that this article can help improve the lavaan software and fix potential conceptual and implementation issues of factor scores in multilevel models.

The article by Finch and French [6], titled "Effect sizes for estimating differential item functioning influence at the test level", presents findings from a simulation study that investigates different effect sizes that can assist in understanding the accumulation of differential item functioning (DIF) at the test score level. The article provides recommendations for the practice of DIF assessment.

The article by Vispoel, Lee, Chen, and Hong [7], titled "Using structural equation modeling to reproduce and extend ANOVA-based generalizability theory analyses for psychological assessments", discusses the estimation of statistical models in generalizability

theory (GT) using structural equation models (SEM). It compares SEM-based estimation with estimation based on analysis of variance (ANOVA) models. The authors provide guidelines for applying the estimation techniques in open-source statistical software. Unfortunately, GT models are less frequently utilized by researchers to assess the reliability of measurement instruments than factor-based models.

The article by Rusch, Venturo-Conerly, Baja, and Mair [8], titled "COPS in action: Exploring structure in the usage of the youth psychotherapy MATCH", is an introduction to the cluster optimized proximity scaling (COPS) method, which is a variant of multidimensional scaling. At the same time, the article serves as a tutorial for the R package COPS that implements the proposed statistical method.

The article by Sorrel, Escudero, Nájera, Kreitchmann, and Vázquez-Lira [9], titled "Exploring approaches for estimating parameters in cognitive diagnosis models with small sample sizes", compares different estimation methods for cognitive diagnostic models (CDM, also referred to as diagnostic classification models (DCM)) in small samples. The study found that alternative estimation methods should be preferred over the usually employed marginal maximum likelihood (MML) estimation approach when estimating CDMs in small samples.

The article by Partchev, Koops, Bechger, Feskens, and Maris [10], titled "dexter: An R package to manage and analyze test data", introduces the R package dexter, which is a professional tool for data management and data analysis in educational assessment programs and survey research. The dexter package focuses on psychometric models that are based on the sum score as sufficient statistics for model parameters, and fits the current data analysis paradigm in the R tidyverse community.

The article by Feuerstahler [11], titled "Applications and extensions of metric stability analysis", describes how to assess metric stability analysis (MSA) for item response theory (IRT) models involving dichotomous and polytomous data. MSA evaluates the effects of item standard errors of item parameters that quantifies how well an IRT model determines the metric of the latent trait.

The article by Karch [12], titled "bmtest: A Jamovi module for Brunner-Munzel's test–A robust alternative to Wilcoxon-Mann-Whitney's test", discusses a Jamovi (a GUI application based on top of the R software) implementation of the Brunner Munzel test for a nonparametric comparison of the locations of two groups. Notably, the Brunner Munzel test overcomes the several limitations of the frequently employed Wilcoxon–Mann–Whitney test.

The article by Luo, De Carolis, Zeng, and Jeon [13], titled "Bayesian estimation of latent space item response models with JAGS, Stan, and NIMBLE in R", serves as a tutorial to estimate the latent space item response model (LSIRM) with the Bayesian general-purpose software packages JAGS, Stan, and NIMBLE. The article contains information on the estimation, specification, convergence assessment, fit evaluation, and result visualization of the LSIRM.

The article by Cole and Paek [14], titled "SAS PROC IRT and the R mirt package: A comparison of model parameter estimation for multidimensional IRT models", compares the estimation performance of multidimensional IRT models in the SAS PROC IRT module and the R package mirt. Both software packages produced nearly identical results in simple loading structure models. In multidimensional IRT models with a complex loading structure or bifactor models, mirt outperformed SAS PROC IRT.

The article by Debelak, Appelbaum, Debeer, and Tomasik [15], titled "Detecting differential item functioning in 2PL multistage assessments", evaluates five different methods for the assessment of DIF in the two-parameter logistic (2PL) IRT model in multi-stage large-scale assessment tests through a simulation study. The type-I error rates of all five approaches were close to the nominal level. A score-based invariance test showed the largest power.

The article by Martinez and Templin [16], titled "Approximate invariance testing in diagnostic classification models in the presence of attribute hierarchies: A Bayesian network

approach", addresses invariance testing regarding items and attribute hierarchies in the log-linear DCM. The invariance testing steps are illustrated using JAGS code.

The article by Zitzmann, Weirich, and Hecht [17], titled "Accurate standard errors in multilevel modeling with heteroscedasticity: A computationally more efficient jackknife technique", compares different standard error estimation methods for multilevel models with heteroscedastic errors. The authors compared the cluster-robust standard error implemented in the Mplus software with the delete-1 jackknife of clusters and a newly proposed method, Zitzmann's jackknife (also known as delete-*d* jackknife). It transpired that Zitzmann's jackknife (slightly) outperformed the other two methods and was computationally much more efficient.

The article by Bulut, Shin, Yildirim-Erbasli, Gorgun, and Pardos [18], titled "An introduction to Bayesian knowledge tracing with pyBKT", provides a tutorial in Bayesian knowledge tracing (BKT). Also, it is a tutorial for estimating BKT models in the pyBKT library in the Python software. Furthermore, different variants of the standard BKT model based on IRT were discussed.

The article by van Erp [19], titled "Bayesian regularized SEM: Current capabilities and constraints", reviews Bayesian regularized SEMs. The author also provides an overview of the various open-source software packages for estimating Bayesian regularized SEMs and illustrates it using an empirical example.

The article by Koopman, Zijlstra, and van der Ark [20], titled "Evaluating model fit in two-level Mokken scale analysis", addresses the assessment of model fit in terms of two-level Mokken scale analysis for clustered data. It is emphasized that the traditional model fit procedure in single-level Mokken scale analysis requires some modifications to adapt it for the two-level case. The proposed method is illustrated in the R package Mokken.

The article by Bailey and Webb [21], titled "Expanding NAEP and TIMSS analysis to include additional variables or a new scoring model using the R package Dire", describes the estimation of conditioning (i.e., latent background) models to draw plausible values (PV) in the R packages, Dire and EdSurvey. The authors highlight that Dire is distinct from other R packages in that it simplifies the evaluation of high-dimensional integrals involved in the estimation of the conditioning model that is required for drawing PVs for composite or subscale variables.

The article by Ye, Kelava, and Noventa [22], titled "Parameter estimation of KST-IRT model under local dependence", discusses the estimation of the KST-IRT model, which combines knowledge space theory (KST) and IRT modeling. In the article, MML estimation and Gibbs sampling were compared through a simulation study.

The article by Kabic and Alexandrowicz [23], titled "RMX/PIccc: An extended person-item map and a unified IRT output for eRm, psychotools, ltm, mirt, and TAM", discusses the implementation of an extension of the well-known person-item map in their R package RMX. The functions in the RMX package cover a wide range of IRT models and are built based on the function output of five frequently used R packages: eRm, ltm, mirt, psychotools, and TAM.

The article by Wagner, Hecht, and Zitzmann [24], titled "A SAS macro for automated stopping of Markov chain Monte Carlo estimation in Bayesian modeling with PROC MCMC", introduces a SAS macro `%automcmc` based on PROC MCMC for the automatic assessment of the convergence of the Markov chain Monte Carlo (MCMC) estimation approach. Users can specify a certain specified stopping criterion for the potential scale reduction factor or the effect sample size, and the SAS macro terminates MCMC estimation if the specified cutoff value is reached.

I would like to thank all the authors of the 23 articles of this Special Issue for their excellent contributions, all of which provided a perfect fit to the scope of the Special Issue. Moreover, I would also like to sincerely thank all the reviewers, editors, and the editorial staff of the *Psych* journal for their support.

Abbreviations

The following abbreviations are used in this manuscript:

2PL	two-parameter logistic
ANOVA	analysis of variance
BKT	Bayesian knowledge tracing
CDM	cognitive diagnostic models
COPS	cluster optimized proximity scaling
DCM	diagnostic classification model
DIF	differential item functioning
DINA	deterministic inputs, noisy "and" gate
GT	generalizability theory
IRT	item response theory
KST	knowledge space theory
LSIRM	latent space item response model
MCMC	Markov chain Monte Carlo
MML	marginal maximum likelihood
MSA	metric stability analysis
PV	plausible value
SEM	structural equation model

References

1. Robitzsch, A. Editorial of the *Psych* special issue "Computational aspects, statistical algorithms and software in psychometrics". *Psych* **2022**, *4*, 114–118. [CrossRef]
2. Dai, S.; Svetina Valdivia, D. Dealing with missing responses in cognitive diagnostic modeling. *Psych* **2022**, *4*, 318–342. [CrossRef]
3. Zitzmann, S.; Walther, J.K.; Hecht, M.; Nagengast, B. What is the maximum likelihood estimate when the initial solution to the optimization problem is inadmissible? The case of negatively estimated variances. *Psych* **2022**, *4*, 343–356. [CrossRef]
4. Sen, S.; Yildirim, I. A tutorial on how to conduct meta-analysis with IBM SPSS statistics. *Psych* **2022**, *4*, 640–667. [CrossRef]
5. Zitzmann, S. A cautionary note regarding multilevel factor score estimates from lavaan. *Psych* **2023**, *5*, 38–49. [CrossRef]
6. Finch, W.H.; French, B.F. Effect sizes for estimating differential item functioning influence at the test level. *Psych* **2023**, *5*, 133–147. [CrossRef]
7. Vispoel, W.P.; Lee, H.; Chen, T.; Hong, H. Using structural equation modeling to reproduce and extend ANOVA-based generalizability theory analyses for psychological assessments. *Psych* **2023**, *5*, 249–273. [CrossRef]
8. Rusch, T.; Venturo-Conerly, K.; Baja, G.; Mair, P. COPS in action: Exploring structure in the usage of the youth psychotherapy MATCH. *Psych* **2023**, *5*, 274–302. [CrossRef]
9. Sorrel, M.A.; Escudero, S.; Nájera, P.; Kreitchmann, R.S.; Vázquez-Lira, R. Exploring approaches for estimating parameters in cognitive diagnosis models with small sample sizes. *Psych* **2023**, *5*, 336–349. [CrossRef]
10. Partchev, I.; Koops, J.; Bechger, T.; Feskens, R.; Maris, G. dexter: An R package to manage and analyze test data. *Psych* **2023**, *5*, 350–375. [CrossRef]
11. Feuerstahler, L. Applications and extensions of metric stability analysis. *Psych* **2023**, *5*, 376–385. [CrossRef]
12. Karch, J.D. bmtest: A Jamovi module for Brunner-Munzel's test—A robust alternative to Wilcoxon-Mann-Whitney's test. *Psych* **2023**, *5*, 386–395. [CrossRef]
13. Luo, J.; De Carolis, L.; Zeng, B.; Jeon, M. Bayesian estimation of latent space item response models with JAGS, Stan, and NIMBLE in R. *Psych* **2023**, *5*, 396–415. [CrossRef]
14. Cole, K.; Paek, I. SAS PROC IRT and the R mirt package: A comparison of model parameter estimation for multidimensional IRT models. *Psych* **2023**, *5*, 416–426. [CrossRef]
15. Debelak, R.; Appelbaum, S.; Debeer, D.; Tomasik, M.J. Detecting differential item functioning in 2PL multistage assessments. *Psych* **2023**, *5*, 461–477. [CrossRef]
16. Martinez, A.J.; Templin, J. Approximate invariance testing in diagnostic classification models in the presence of attribute hierarchies: A Bayesian network approach. *Psych* **2023**, *5*, 688–714. [CrossRef]
17. Zitzmann, S.; Weirich, S.; Hecht, M. Accurate standard errors in multilevel modeling with heteroscedasticity: A computationally more efficient jackknife technique. *Psych* **2023**, *5*, 757–769. [CrossRef]
18. Bulut, O.; Shin, J.; Yildirim-Erbasli, S.N.; Gorgun, G.; Pardos, Z.A. An introduction to Bayesian knowledge tracing with pyBKT. *Psych* **2023**, *5*, 770–786. [CrossRef]
19. van Erp, S. Bayesian regularized SEM: Current capabilities and constraints. *Psych* **2023**, *5*, 814–835. [CrossRef]
20. Koopman, L.; Zijlstra, B.J.H.; van der Ark, L.A. Evaluating model fit in two-level Mokken scale analysis. *Psych* **2023**, *5*, 847–865. [CrossRef]

21. Bailey, P.D.; Webb, B. Expanding NAEP and TIMSS analysis to include additional variables or a new scoring model using the R package Dire. *Psych* **2023**, *5*, 876–895. [CrossRef]

22. Ye, S.; Kelava, A.; Noventa, S. Parameter estimation of KST-IRT model under local dependence. *Psych* **2023**, *5*, 908–927. [CrossRef]

23. Kabic, M.; Alexandrowicz, R.W. RMX/PIccc: An extended person-item map and a unified IRT output for eRm, psychotools, ltm, mirt, and TAM. *Psych* **2023**, *5*, 948–965. [CrossRef]

24. Wagner, W.; Hecht, M.; Zitzmann, S. A SAS macro for automated stopping of Markov chain Monte Carlo estimation in Bayesian modeling with PROC MCMC. *Psych* **2023**, *5*, 966–982. [CrossRef]