



## An Introduction to the Special Volume on “Psychometrics in R”

**Jan de Leeuw**

University of California at Los Angeles

**Patrick Mair**

Wirtschaftsuniversität Wien

---

### Abstract

This special volume presents a select number of psychometric techniques, many of them original, and their implementation in R packages.

*Keywords:* psychometrics, R.

---

## 1. Introduction

This is the second volume in the series “Foometrics in R”. The first volume, edited by Kate Mullen and Ivo van Stokkum, was on chemometrics and spectrometrics. It was published as Volume 18 of the *Journal of Statistical Software* (JSS). As explained in the introduction to that volume, it was decided at the useR! 2006 conference in Vienna to publish a number of these volumes in JSS. Volumes on R ([R Development Core Team 2007](#)) in political methodology, ecology, meta-analysis, social network analysis, and econometrics are in preparation.

For this particular volume, on psychometrics in R, guest editor Jan de Leeuw is a guest in his own house. Guest editor Patrick Mair reviewed all R code, including his own. Jan de Leeuw originally decided to edit this volume because he happened to know quite a few people working in psychometrics, and because he thought the field needed more exposure to R. Early in the history of statistics, there was no distinction between psychometrics and the rest of statistics. Galton, Pearson, and Edgeworth all made important contributions to both intelligence and achievement testing. And it has remained true, throughout the history of psychometrics, that many of the important contributions to the field were made by statisticians such as Hotelling, Kelley, Lawley, Anderson, Rasch, Jöreskog, Kruskal, Benzécri, and so on. Nevertheless, at some point psychometrics and statistics started to diverge. Psychometricians basically created their own version of multivariate analysis, their own version of experimental design and fixed effects analysis of variance, and their own version of null hypothesis significance testing. And

these localized versions of statistical rituals developed their own languages, both ordinary and computer languages. Psychometricians developed their own software packages, and in fact SPSS, SAS, SYSTAT, Minitab, DataDesk, and JMP all have strong to moderately strong historical connections with psychometrics.

For quite a long time the communication between the fields basically stopped, except in some centers such as ETS and Bell Labs. Academic statisticians limited their conversations with psychometricians to snotty remarks about factor analysis, while psychometricians joked about the gothic irrelevances erected in the *Annals of Mathematical Statistics*. This was, of course, unfortunate. And one of the unfortunate consequences is that R, the *lingua franca* of computational statistics, has not been used much in psychometrics. It is my impression that MATLAB is more popular, which is somewhat unfortunate, because MATLAB is not open source and has far fewer statistics add-ons.

One would have hoped that the increasing emphasis on data and computation, as well as the slow demise of both the factor analysis and the mathematical statistics cults, would have brought the two disciplines closer together. So far this has not really happened. Hopefully, this volume will at least show psychometricians there are advantages in working with R, advantages that have to do with availability of many tested and supported high-level packages, and advantages that have to do with being integrated into a much larger community of programmers and scholars.

## 2. An outline of the contributions

As is the case in psychometrics proper, the emphasis in this volume is on test theory, in particular on item response theory (IRT). In fact, the volume may have a little bit too much emphasis on IRT and the Rasch model, but that is partly because it was easier to solicit contributions from people working in that field.

The first paper by [Doran, Bates, Bliese, and Dowling \(2007\)](#) combines two of the most active areas in psychometrics. The Rasch model is extended to the multilevel case, to a mixed Rasch model, using Laplace approximations for the mixed model likelihood. It is then shown how the `lmer()` function in the tried and tested **lme4** package can be used to fit such models.

There are many ways to perform a correspondence analysis in R. For instance, one can use the recommended package **MASS** ([Venables and Ripley 2002](#)) or the contributed package **ade4** ([Chessel, Dufour, and Thioulouse 2004](#)). [Nenadić and Greenacre \(2007\)](#) provide a specialized package **ca**, which does simple, multiple, and joint correspondence analysis with supplementary points and an abundance of possible graphical displays.

Back to Rasch. [Verhelst, Mair, and Hatzinger \(2007\)](#) present a specialized MCMC method to sample from the set of binary matrices with given marginals. The Rasch model implies that, conditional on the marginals, all binary matrices have equal probability. Thus the **RaschSampler** package described in this paper can be used for non-parametric testing of all possible statistics on the space of binary matrices.

There is more IRT, MCMC, and multilevel in the next contribution by [Fox \(2007\)](#). The package **mlirt** implements an MCMC method to fit a multilevel version of the ordinal probit model for polytomous items. The model is embedded in a Bayesian framework with the usual conjugate priors, and a combination of data augmentation and Gibbs sampling is used to construct the MCMC algorithm.

Anderson, Li, and Vermunt (2007) present the **plRasch** package. This interprets the Rasch model as a log-linear-by-linear association (LLLA) model, which leads to natural extensions to polytomous data and multiple latent traits. Exact likelihood estimation (at least by deterministic methods) is not feasible, but Anderson et al. show that pseudo-likelihood methods perform well and lead to little loss of efficiency.

Fox, Klein Entink, and van der Linden (2007) apply IRT to modeling both responses and response times. As they point out, response times are readily available in computerized testing. The joint likelihood is build up hierarchically, using the two-parameter normal ogive model for the responses and a log-normal model for the response times. MCMC methods are implemented to estimate the conjoint IRT model in the package **cirt**.

After the avalanche of IRT and MCMC, Kelley (2007) presents the **mbess** package to construct confidence intervals for standardized effect sizes. There is a tendency to replace null hypothesis significance testing, with its binary decisions, by the far more informative use of effect size. There are many statistics in use for measuring effects, and Kelley's package uses the non-central distributions available in R to compute confidence intervals for them.

More Rasch. The **eRm** package of Mair and Hatzinger (2007) fits various extensions of the Rasch model, in particular rating scale models and partial credit models, as well as various unidimensional linear logistic test models. Throughout, conditional maximum likelihood methods are used, which allow for separation of the item and person parameters. The package emphasizes ease of use and graphical fit diagnostics.

Johnson (2007) generalizes the logistic partial credit model in his **gpcm** package. He has not yielded to the temptation to go Bayesian and using MCMC. **gpcm** uses Gauss-Hermite quadrature and marginal maximum likelihood estimation using the EM algorithm. It can be thought of as an extension, in one direction, of the **ltm** package that has been available on CRAN for some time.

In the final contribution to the volume van der Ark (2007) discusses the **mokken** package for non-parametric item analysis. Mokken scale analysis is a non-parametric form of IRT, developed more than 30 years ago. It has been used mostly in the Netherlands, and most of the theory and the software has also been developed there. Van der Ark's paper, and the corresponding R package, can make this interesting approach to IRT more widely known and more easily used.

### 3. Conclusions and outlook

As we said before, IRT is certainly not underrepresented in this special volume. There is relatively little material on factor analysis and structural equation modeling of continuous variables, on multidimensional scaling, and on optimal scaling techniques. And these are all important components of psychometrics. Instead of worrying about this possible lack of representativeness, however, what we should do (and will do) is to prepare a second volume along the same lines, concentrating more on the analysis of multivariate data with continuous observed variables. If you have suggestions, or even a contribution, contact the guest editor.

In all other respects, we think, this volume does the job we intended it to do. We see the versatility of R, and the use of built-in functions and previously existing packages to rapidly create new methods. And the volume adds at least eleven easily accessible and interesting techniques to the toolkit of any psychometrician or applied statistician.

## References

- Anderson CJ, Li Z, Vermunt JK (2007). “Estimation of Models in a Rasch Family for Polytomous Items and Multiple Latent Variables.” *Journal of Statistical Software*, **20**(6). URL <http://www.jstatsoft.org/v20/i06/>.
- Chessel D, Dufour AB, Thioulouse J (2004). “The **ade4** Package – I: One-table Methods.” *R News*, **4**(1), 5–10.
- Doran H, Bates D, Bliese P, Dowling M (2007). “Estimating the Multilevel Rasch Model: With the **lme4** Package.” *Journal of Statistical Software*, **20**(2). URL <http://www.jstatsoft.org/v20/i02/>.
- Fox JP (2007). “Multilevel IRT Modeling in Practice with the Package **mlirt**.” *Journal of Statistical Software*, **20**(5). URL <http://www.jstatsoft.org/v20/i05/>.
- Fox JP, Klein Entink R, van der Linden W (2007). “Modeling of Responses and Response Times with the Package **cirt**.” *Journal of Statistical Software*, **20**(7). URL <http://www.jstatsoft.org/v20/i07/>.
- Johnson M (2007). “Marginal Maximum Likelihood Estimation of Item Response Models in R.” *Journal of Statistical Software*, **20**(10). URL <http://www.jstatsoft.org/v20/i10/>.
- Kelley K (2007). “Constructing Confidence Intervals for Standardized Effect Sizes: Theory, Application, and Implementation.” *Journal of Statistical Software*, **20**(8). URL <http://www.jstatsoft.org/v20/i08/>.
- Mair P, Hatzinger R (2007). “Extended Rasch Modeling: The **eRm** Package for the Application of IRT Models in R.” *Journal of Statistical Software*, **20**(9). URL <http://www.jstatsoft.org/v20/i09/>.
- Nenadić O, Greenacre M (2007). “Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The **ca** Package.” *Journal of Statistical Software*, **20**(3). URL <http://www.jstatsoft.org/v20/i03/>.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- van der Ark LA (2007). “Mokken Scale Analysis in R.” *Journal of Statistical Software*, **20**(11). URL <http://www.jstatsoft.org/v20/i11/>.
- Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. Springer-Verlag, New York, fourth edition. ISBN 0-387-95457-0.
- Verhelst N, Mair P, Hatzinger R (2007). “The Rasch Sampler.” *Journal of Statistical Software*, **20**(4). URL <http://www.jstatsoft.org/v20/i04/>.

**Affiliation:**

Jan de Leeuw  
Department of Statistics  
University of California at Los Angeles  
Box 951554  
Los Angeles, CA 90095-1554, United States of America  
E-mail: [deleeuw@stat.ucla.edu](mailto:deleeuw@stat.ucla.edu)  
URL: <http://www.cuddyvalley.org/>