Book Review

Applied Psychological Measurement
2018, Vol. 42(7) 590–592
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0146621617752992
journals.sagepub.com/home/apm

Jorge González and Marie Wiberg

Applying Test Equating Methods: Using R. New York, NY: Springer, 2017. 196 pp. Hardcover price \$99.99, eBook price \$69.99. Hardcover ISBN: 978-3-319-51822-0, eBook ISBN: 978-3-319-51824-4

Reviewed by: Michela Battauz , Department of Economics and Statistics, University of Udine, Italy

Testing is used in many countries for many different purposes, including assessing academic achievement, making higher education admissions decisions, and certifying language proficiency. Test security requires the use of different versions of a test over different administrations and results in scores that are not directly comparable. The purpose of the equating process is to make these scores comparable, and thus constitutes a fundamental step in providing adequate information to decision makers while maintaining a fair assessment of examinees.

The publication of books on the theory of test equating (see Kolen & Brennan, 2014; von Davier, 2011; von Davier, Holland, & Thayer, 2004) has been fundamental, in terms of disseminating the methodology and *Applying Test Equating Methods: Using R* fills the existing void between theory and practice. The book explains the process of test equating from a theoretical perspective, as well as an applied perspective, using the software R (R Development Core Team, 2017). R is a free and open source software that is widely used for statistical analyses in both academia and the private sector. The great strength of R is that users can contribute additional packages, making them available for the entire R community. Consequently, the number of packages available on the public repository has grown exponentially over time, thus providing a way for users to utilize a large variety of statistical, psychometric, and graphical, procedures. This has resulted in the development of a large number of packages with partially overlapping functionalities and different structures. To address this issue, books on the use of R for a particular field are much appreciated by the community of researchers who use R. This book is a welcome addition to this arena, considering all of the available packages in R that are available for equating, as it provides a clear description of the functionalities of each of the packages, with respect to the underlying theories. After providing a general theoretical background of equating, every chapter of the book is dedicated to a class of methods, which are first explained theoretically and then applied using R.

Chapter 1 provides the statistical framework of test equating, describes different data collection designs, and presents the R packages described within the book. The main packages described in the book are *equate* (Albano, 2016), *SNSequate* (González, 2014), *kequate* (Andersson, Bränberg, & Wiberg, 2013), and *equateIRT* (Battauz, 2015).

Chapter 2 introduces the data used for analyses described in the book, including data available in the books written by Kolen and Brennan (2014) and von Davier et al. (2004), as well as data collected for assessing academic achievement and making college admissions decisions. All the data are available in R packages or on the book's webpage, thus allowing the interested reader to run their own analyses using these datasets. Chapter 2 also explains how to create and

visualize frequency distributions for overall scores that correspond to different data collection designs. Finally, Chapter 2 describes the methodology used to obtain a smooth distribution of scores. This procedure is known as presmoothing, and constitutes the first step of several equipercentile equating methods.

Chapter 3 describes traditional equating methods, which include equipercentile equating, linear equating, and mean equating. Focusing in particular on the NEAT design, which is the more challenging equating scenario, in that it requires the choice of specific assumptions, this chapter describes various equating methods to perform equipercentile or linear equating. Examples in this chapter include all the data collection designs and utilize the R packages *equate* and *SNSequate*.

Kernel equating is presented in Chapter 4. After describing presmoothing and estimation of score probabilities, the authors explain the kernel method of continuization of the discrete score distribution. Finally, this chapter treats the computation of equated scores and accuracy measures. Examples of all the phases of kernel equating are provided, using the packages *SNSequate* and *kequate*.

Item response theory (IRT) equating is illustrated in Chapter 5. After a brief review of IRT models for binary responses, this chapter explains available methods for converting item parameters and ability values to a common metric, and describes the procedures for obtaining equated scores using an IRT approach. The application of these methods to real data examples is illustrated using the R packages *ltm* (Rizopoulos, 2006) and *mirt* (Chalmers, 2012) for the estimation of the IRT models, while the packages *SNSequate* and *equateIRT* are used for equating. Finally, concurrent calibration and fixed item parameter calibration are also considered in this chapter.

Chapter 6 focuses on the recently proposed method of local equating, which considers the conditional distributions of test scores given ability level, as opposed to the marginal distributions. After presenting the general idea underlying local equating, the authors describe various methods that utilize this approach. These methods include both traditional and IRT techniques. The R packages used to implement these methods are *kequate* and *SNSequate*.

Recent developments in equating are considered in Chapter 7. Within the kernel equating framework, these newly proposed methodologies include the use of two alternative kernels, two novel methods for bandwidth selection, and IRT kernel equating for both binary and polytomous items. Furthermore, this chapter presents a Bayesian nonparametric model for equating, and a method for assessing equating transformations obtained with different methods. Examples are provided using the packages *SNSequate* and *kequate*, which implement these new methods.

Two appendixes are included at the end of the book. The first appendix gives instructions for installing R and additional packages, and for importing data of different file formats. The second appendix contains some technical details related to the methods presented throughout the book.

Overall, my impression of the book is extremely positive as the authors achieved the goal of providing a clear and practical guide that explains which packages implement each method, coupled with reproducible examples. The only thing lacking is that it would have been of interest to add a discussion on which relevant methods for test equating (if any) are not yet implemented in R, to guide who is interested in software development to develop these missing procedures.

This book can be of interest to both psychometricians and statisticians interested in equating. It can be used as a concise theoretical manual on test equating, but can also be used to guide practical implementation of the theoretical methods. For these reasons, the book is suited for graduate and postgraduate students as well as more experienced researcher and practitioners.

ORCID iD

Michela Battauz <https://orcid.org/0000-0002-3098-689X>

References

- Albano, A. D. (2016). equate: An R package for observed-score linking and equating. *Journal of Statistical Software*, 74(8), 1-36.
- Andersson, B., Bränberg, K., & Wiberg, M. (2013). Performing the kernel method of test equating with the package kequate. *Journal of Statistical Software*, 55(6), 1-25.
- Battauz, M. (2015). equateIRT: An R package for IRT test equating. *Journal of Statistical Software*, 68(7), 1-22.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29.
- González, J. (2014). SNSequate: Standard and nonstandard statistical models and methods for test equating. *Journal of Statistical Software*, 59(7), 1-30.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling and linking: Methods and practices* (3rd ed.). New York, NY: Springer.
- R Development Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: The R Foundation for Statistical Computing.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1-25.
- von Davier, A. A. (Ed.). (2011). *Statistical models for test equating, scaling, and linking*. New York, NY: Springer.
- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The kernel method of test equating*. New York, NY: Springer.