SEAsic: R Package for Score Equity Assessment

Applied Psychological Measurement
2015, Vol. 39(6) 496–498
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146621615572832
apm.sagepub.com

Anne Corinne Huggins-Manley¹ and
Douglas Whitaker¹

Abstract

SEAsic (score equity assessment–summary index computation) is an R package for computing and graphing a variety of indices that quantify an important aspect of test fairness, that of reported score equity. Historically, test fairness has been statistically defined as a lack of differential predication and/or a presence of measurement invariance at the item level. More recent definitions of fairness include the concept of score equity, which calls for additional subpopulation analysis at the equated, reported test score level. SEAsic allows for efficient calculation and graphing of multiple score equity assessment (SEA) indices. All indices in Huggins and Penfield (2012) can be calculated and plotted in various ways given a user-provided conversion table. SEAsic is freely available on the Comprehensive R Archive Network (CRAN). Mac, Windows, and Linux users have access to the package via downloading the appropriate version of R or RStudio from the CRAN website. Multiple examples of each index computation, variations on each index, and plot options are provided in the package manual on CRAN.

Keywords

score equity, equating invariance, test fairness, R package

SEAsic (score equity assessment–summary index computation) is an R package for computing and graphing a variety of indices that quantify an important aspect of test fairness, that of reported score equity. Historically, test fairness has been statistically defined as a lack of differential predication (e.g., Linn, 1982) and a presence of measurement invariance at the item level (e.g., Zumbo, 1999). More recent definitions of fairness include the concept of score equity, which calls for additional subpopulation analysis at the equated and reported test score level (Dorans, 2004b). The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) emphasize fairness as a lack of bias with specific reference to the reported scores for examinees from different subpopulations. The Joint Committee on Testing Practices (2004) and the U.S. Department of Education (2009) mandate that test developers ensure comparability of reported mean scores that are disaggregated at the subpopulation level. Score equity assessment (SEA; Dorans, 2004b) is aimed at quantifying this type of score fairness.

¹University of Florida, Gainesville, FL, USA

Corresponding Author:

Anne Corinne Huggins-Manley, Research and Evaluation Methodology, University of Florida, 1215 Norman Hall,
Gainesville, FL 32611, USA.
Email: ahuggins@coe.ufl.edu

SEA is playing an increasing role in the fairness literature (e.g., special issues of *Applied Psychological Measurement*, “Population Invariance,” 2008, and *Journal of Educational Measurement*, Dorans, 2004a). Many SEA indices have been proposed. Some quantify the degree to which score equity holds across the score range of a test (i.e., unconditional indices) while others isolate this quantification at individual score levels (i.e., conditional indices). Some indices compare all relevant subpopulations to a composite population (i.e., omnibus indices), others compare single subpopulations to a composite population (i.e., group-to-overall indices), and others directly compare pairs of subpopulations (i.e., pairwise indices). The majority of SEA applications focus on a single or select few indices while ignoring others (e.g., Liu, Cahn, & Dorans, 2006; Yang, 2004; Yi, Harris, & Gao, 2008). However, SEA application can be improved by including multiple indices of test score fairness.

SEAsic allows for efficient calculation and graphing of multiple SEA indices. All indices in Huggins and Penfield (2012) can be calculated and plotted in various ways given a user-provided conversion table. Included are the root expected mean square difference (REMSD; Dorans & Holland, 2000), its group-to-overall counterpart (Yang, 2004), and two conditional counterparts, as well as options for manipulating subpopulation and/or score level weights in all of these indices (see Brennan, 2008). All indices from Kolen and Brennan (2004) are included, such as the pairwise mean difference (MD) and absolute mean difference (MAD) indices with options for equal weights (i.e., *ewMD*, *ewMAD*) and conditional counterparts. Multiple indices that locate maximum differences (Dorans & Holland, 2000) are also included. Every index in SEAsic can be computed and plotted in its unstandardized or standardized form, and the option of comparing rounded or unrounded equivalents (see Brennan, 2008) is left to the user in her specification of the conversion table. All index results can be plotted alone or combined into plots that provide a variety of visual summaries of score equity, including a user-specified reference line at the (un)standardized difference that matters (Dorans & Feigenbaum, 1994).

SEAsic is freely available on the Comprehensive R Archive Network (CRAN) at <http://www.cran.r-project.org> or can be installed through R (R Development Core Team, 2005) or RStudio (2012) with `install.packages("SEAsic")`. Multiple examples of each index computation, variations on each index, and plot options are provided in the package manual on CRAN. Mac, Windows, and Linux users have access to the package via downloading the appropriate version of R or RStudio from the CRAN website. Documentation of the package is also freely available on CRAN and conforms to the requisite documentation for all R packages.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Brennan, R. L. (2008). A discussion of population invariance. *Applied Psychological Measurement*, 32, 102-114.

- Dorans, N. J. (Ed.). (2004a). Assessing the population sensitivity of equating functions [Special issue]. *Journal of Education Measurement*, 41, 1-81.
- Dorans, N. J. (2004b). Using subpopulation invariance to assess test score equity. *Journal of Educational Measurement*, 41, 43-68.
- Dorans, N. J., & Feigenbaum, M. D. (1994). Equating issues engendered by changes to the SAT and PSAT/NMSQT. In I. M. Lawrence, N. J. Dorans, M. D. Feigenbaum, N. J. Feryok, A. P. Schmitt, & N. K. Wright (Eds.), *Technical issues related to the introduction of the new SAT and PSAT/NMSQT* (ETS Research Memorandum No. RM-94-10). Princeton, NJ: Educational Testing Service.
- Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equitability of tests: Theory and the linear case. *Journal of Educational Measurement*, 37, 281-306.
- Huggins, A. C., & Penfield, R. D. (2012). An NCME instructional module on population invariance in linking and equating. *Educational Measurement: Issues and Practices*, 31, 27-40.
- Joint Committee on Testing Practices. (2004). *Code for fair testing practices in education*. Washington, DC: Author. Retrieved from <http://www.apa.org/science/programs/testing/fair-code.aspx>
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer.
- Linn, R. L. (1982). Ability testing: Individual differences and differential prediction. In A. K. Wigdor & W. R. Garner (Eds.), *Ability testing: Uses, consequences, and controversies, Part II* (pp. 335-388). Washington, DC: National Academy Press.
- Liu, J., Cahn, M. F., & Dorans, N. J. (2006). An application of score equity assessment: Invariance of linkage of new SAT® to old SAT across gender groups. *Journal of Educational Measurement*, 43, 113-129.
- Population invariance [Special issue]. (2008). *Applied Psychological Measurement*, 32, 9-114.
- R Development Core Team. (2005). *R: A language and environment for statistical computing, reference index version 2.2.1*. Vienna, Austria: R Foundation for Statistical Computing. Available from <http://www.R-project.org>
- RStudio. (2012). RStudio: Integrated development environment for R (Version 0.96.122) [Computer software]. Boston, MA. Available from <http://www.rstudio.org/>
- U.S. Department of Education. (2009). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Washington, DC: Author.
- Yang, W. L. (2004). Sensitivity of linkings between AP multiple-choice scores and composite scores to geographical region: An illustration of checking for population invariance. *Journal of Educational Measurement*, 41, 33-41.
- Yi, Q., Harris, D.J., & Gao, X. (2008). Invariance of equating functions across different subgroups of examinees taking a science achievement test. *Applied Psychological Measurement*, 32, 62-80.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF)*. Ottawa, Ontario, Canada: National Defense Headquarters.