MCMC Z-G: An IRT Computer Program for Forced-Choice Noncognitive Measurement

Applied Psychological Measurement
2016, Vol. 40(7) 551–553
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146621616663682
apm.sagepub.com

Wei Wang¹, Philseok Lee², Seang-Hwane Joo²,
Stephen Stark², and Robert Louden¹

Abstract

In recent years, there has been a surge of interest in measuring noncognitive constructs in educational and managerial/organizational settings. For the most part, these noncognitive constructs have been and continue to be measured using Likert-type (ordinal response) scales, which are susceptible to several types of response distortion. To deal with these response biases, researchers have proposed using forced-choice format, which requires respondents or raters to evaluate cognitive, affective, or behavioral descriptors presented in blocks of two or more. The workhorse for this measurement endeavor is the item response theory (IRT) model developed by Zinnes and Griggs (Z-G), which was first used as the basis for a computerized adaptive rating scale (CARS), and then extended by many organizational scientists. However, applications of the Z-G model outside of organizational contexts have been limited, primarily due to the lack of publicly available software for parameter estimation. This research effort addressed that need by developing a Markov chain Monte Carlo (MCMC) estimation program, called MCMC Z-G, which uses a Metropolis-Hastings-within-Gibbs algorithm to simultaneously estimate Z-G item and person parameters. This publicly available computer program MCMC Z-G can run on both Mac OS® and Windows® platforms.

Keywords

item response theory, Zinnes–Griggs or Z-G model, forced-choice, noncognitive measurement, MCMC

In recent years, there has been a surge of interest in measuring noncognitive constructs in educational and managerial/organizational settings. Examples include personality, vocational interests,

---

¹University of Central Florida, Orlando, USA

²University of South Florida, Tampa, USA

Corresponding Authors:

Stephen Stark, Department of Psychology, University of South Florida, 4202 E. Fowler Ave., PCD 4118G, Tampa, FL 33620, USA.

Email: sestark@usf.edu

Wei Wang, Department of Psychology, University of Central Florida, 4111 Pictor Lane, Orlando, FL 32816, USA.

Email: wei.wang@ucf.edu

motivation, job satisfaction, job performance, and person-environment fit. For the most part, these noncognitive constructs have been and continue to be measured using Likert-type (ordinal response) scales, which are susceptible to several types of response distortion (Brown & Maydeu-Olivares, 2011). In personnel selection, a key concern is faking good, also known as impression management or socially desirable responding (Stark, Chernyshenko, & Drasgow, 2005). In student or employee performance appraisal, a primary concern is rating scale errors, which include severity, leniency, and central tendency errors (Borman et al., 2001), that tend to diminish variance in scores and undermine decision-making and developmental feedback efforts.

To deal with these response biases, researchers have proposed using forced-choice format, which requires respondents or raters to evaluate cognitive, affective or behavioral descriptors presented in blocks of two or more. The respondent's task is to choose the descriptor in each block that better characterizes himself or herself, or, alternatively a ratee who is the external target of the evaluation (Stark & Drasgow, 2002).

Borman and colleagues (2001) proposed using a unidimensional forced-choice item response theory (IRT) model, developed by Zinnes and Griggs (Z-G; 1974), as the basis for a computerized adaptive employee appraisal system known as CARS II. To support this application, Stark and Drasgow (2002) developed statement and person estimation methods for the Z-G model, as well as a computerized adaptive testing algorithm (Stark & Chernyshenko, 2011; Stark & Drasgow, 1998) to enhance measurement precision. Since then, the Z-G model has been used for some large personality testing applications in civilian and military settings (e.g., the U.S. Navy Computerized Adaptive Personality Scales [NCAPS]; Houston, Borman, Farmer, & Bearden, 2006; Stark et al., 2014) and, conceptually, it is suitable for assessing a wide array of noncognitive constructs involving "self-" and "other" reports. Applications of the Z-G model outside of organizational contexts have been limited, however, by the lack of publicly available software for parameter estimation. This research effort addressed that need by developing a Markov chain Monte Carlo (MCMC) estimation program, called MCMC Z-G, which uses a Metropolis-Hastings-within-Gibbs algorithm (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) to simultaneously estimate Z-G item and person parameters (Seybert, Lee, Stark, & Chernyshenko, 2014). The MCMC Z-G program can handle missing data, and there is no limit on the number of items or respondents.

Like the MCMC GGUM (Generalized Graded Unfolding Model; Roberts, Donoghue, & Laughlin, 2000) program developed by Wang, de la Torre, and Drasgow (2014), the MCMC Z-G program is run by a syntax file, which specifies data characteristics (e.g., file location, pair mapping, initial values, number of respondents, items, statements), MCMC estimation specifications (e.g., number of chains, chain length, and burn-in iterations), and the standard deviations for sampling candidates. The output file includes the means and standard errors of the item and person parameters from each estimation chain, as well as the averaged values across the chains. Acceptance rates are reported for item parameter estimation.

The executable MCMC Z-G installer is a C++ application having a user-friendly graphical interface that will run on both Mac OS® and Windows® platforms. The installers and corresponding documentation, such as exemplary syntax files and data sets, are available from <http://computationalpsychology.org/resources/>. It is free of charge for noncommercial use only.

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

- Borman, W. C., Buck, D., Hanson, M. A., Motowidlo, S. J., Stark, S., & Drasgow, F. (2001). An examination of the comparative reliability, validity, and accuracy of performance ratings made using computerized adaptive rating scales. *Journal of Applied Psychology, 86*, 965-973.
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71*, 460-502.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika, 57*, 97-109.
- Houston, J. S., Borman, W. C., Farmer, W. L., & Bearden, R. M. (2006). *Development of the Navy Computer Adaptive Personality Scales (NCAPS)* (NPRST-TR-06-2). Millington, TN: Navy Personnel Research, Studies, and Technology Division, Bureau of Naval Personnel.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics, 21*, 1087-1092.
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*, 3-32.
- Seybert, J., Lee, P., Stark, S., & Chernyshenko, O. S. (2014). *Evaluating methods for pairwise preference item calibration under the Zinnes-Griggs item response theory model*. Unpublished manuscript.
- Stark, S., & Chernyshenko, O. S. (2011). Computerized adaptive testing with the Zinnes and Griggs pairwise preference ideal point model. *International Journal of Testing, 11*, 231-247. doi:10.1080/15305058.2011.561459
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise preference model. *Applied Psychological Measurement, 29*, 184-201.
- Stark, S., Chernyshenko, O. S., Drasgow, F., White, L. A., Heffner, T., Nye, C. D., & Farmer, W. L. (2014). From ABLE to TAPAS: A new generation of personality tests to support military selection and classification decisions. *Military Psychology, 26*, 153-164.
- Stark, S., & Drasgow, F. (1998, April). *Application of an item response theory ideal point model to computer adaptive assessment of job performance*. Paper presented at the 13th annual conference for the Society of Industrial and Organizational Psychology, Dallas, TX.
- Stark, S., & Drasgow, F. (2002). An EM approach to parameter estimation for the Zinnes and Griggs paired comparison IRT model. *Applied Psychological Measurement, 26*, 208-227.
- Wang, W., de la Torre, J., & Drasgow, F. (2015). MCMC GGUM: A new computer program for estimating unfolding IRT models. *Applied Psychological Measurement, 39*, 160-161. doi:10.1177/0146621614540514
- Zinnes, J. L., & Griggs, R. A. (1974). Probabilistic, multidimensional unfolding analysis. *Psychometrika, 39*, 327-350.