A Review of SimuMCAT: A Simulation Software for Multidimensional Computerized Adaptive Testing

Applied Psychological Measurement

2015, Vol. 39(3) 241–244

© The Author(s) 2015

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0146621615569503

apm.sagepub.com

Haskell Sie¹

Abstract

This article reviews the software package SimuMCAT that simulates unidimensional and multidimensional computerized adaptive testing with various types of items (dichotomous/polytomous) and loading structures (simple-/complex-structured). In addition, the software allows users to choose from five different item selection procedures, two stopping rules for variable-length tests, as well as test constraints to satisfy test blueprint and limit item exposure.

Keywords

item response theory, computerized adaptive testing, exposure control, content balancing, stopping rule

Computerized adaptive testing (CAT) has long been discussed in the psychometric literature, with applications ranging from educational testing to health and psychological assessments. With an increased interest in multidimensional item response theory (MIRT) models that facilitate subscores reporting in addition to one overall score, a natural direction in the CAT literature has been to extend the benefits of unidimensional CAT to multidimensional CAT (MCAT). Compared with its unidimensional counterpart, MCAT leads to higher precision and reliability as well as reduces test length (Segall, 1996; W. C. Wang & Chen, 2004). Applications of MCAT have been discussed in the literature, both to estimate test takers' latent abilities (Veldkamp & van der Linden, 2002; Yao, Pommerich, & Segall, 2013) as well as to make a classification decision (Luecht, 1996). Such applications of MCAT have been supported by advancements in research studies pertaining to its administration, including comparison between item selection methods (Mulder & van der Linden, 2009; C. Wang & Chang, 2011; C. Wang, Chang, & Boughton, 2011; Yao, 2012), construction of an optimal item pool (Yao, 2014), as well as incorporation of stopping rules for variable-length MCAT (Yao, 2013). In addition, MCAT procedures that can incorporate both multiple-choice (MC) and constructed-response (CR) items have also been developed for assessments consisting of a mixture of those item types (Yao & Schwarz, 2006).

¹American Institutes for Research, Washington, DC, USA

Corresponding Author:

Haskell Sie, American Institutes for Research, 1000 Thomas Jefferson Street NW, Washington, DC 20007-3835, USA.

Email: hsie@air.org

Currently, there are only a few software programs with a built-in capacity to administer MCATs, especially when additional constraints such as content balancing and item exposure control are used. This is primarily due to the complexity of MIRT models and the computational burden often encountered when selecting in real-time multidimensional test items to match examinees' multidimensional abilities. This article focuses on the software program SimuMCAT (Yao, 2011), which has comprehensive capabilities to administer MCAT. For more general purposes, including multidimensional calibration and multidimensional linking prior to treating the item parameters as known constants when administering MCAT, the software can be used in conjunction with BMIRT (Yao, 2003) and LinkMIRT (Yao, 2004).

Obtaining the Software, Installation, and Documentation

SimuMCAT is non-commercial software that can be downloaded for free from www.BMIRT.com, along with BMIRT and LinkMIRT. To use this software, the user must have the Java Runtime Environment (JRE) installed, which can be downloaded from <https://www.java.com/en/download/>. The user needs to ensure that JRE is installed in the proper location on the computer.

Program documentation for BMIRT, LinkMIRT, and simuMIRT can be found in a single user manual. Chapter 6 of the user manual includes technical details of the five adaptive item selection methods supported by the software as well as two stopping rules that users can choose from when simulating a variable-length MCAT. Chapter 7 includes descriptions on the two input files and the two output files. A list of the functions needed to implement combinations of the item selection method, the content constraint, the exposure control procedure, and the stopping rules for variable-length tests is also included in Chapter 7.

Input and Output Files

Examples of input files are included with the initial download of SimuMCAT. The examples as well as the description of how to prepare an input file given in the user manual are relatively straightforward for a new user of SimuMCAT to follow. The user is also given flexibility to choose other values for the simulation conditions in lieu of the default. Output files of SimuMCAT are clearly labeled. The .ss output file has one line for each examinee, with information about his or her true abilities (domain and overall), estimated abilities (domain and overall), standard errors of estimated abilities, testing time, test reliability, as well as final optimized angles to compute the optimized overall score. More detailed information about the test administered for each examinee is available in the .par output file. This file contains the ID as well as the parameters of each item administered to each examinee. It also contains the examinee's scored response and his or her updated domain ability estimates and the corresponding standard errors.

Program Functionalities

SimuMCAT facilitates CAT simulation without any practical restrictions on the number of examinees, the number of items in the pool, or the number of dimensions. Different types of items are allowed in the item pool, including dichotomous items modeled using the multidimensional three-parameter logistic model (M3PL; Reckase, 1997) and polytomous items modeled using the multidimensional two-parameter partial credit model (M2PPC; Yao & Schwarz, 2006). In addition, the item pool can also contain a mixture of simple- and complex-structured items.

SimuMCAT supports several different multidimensional item selection methods that go beyond the standard practice based on Fisher and Kullback–Leibler information. In particular, three new methods to choose test items are implemented, based on either minimizing variance of an overall score estimated using optimized weights of the domain scores or maximizing information in the direction currently lacking information (i.e., a maximin principle). The item selection methods can be implemented in conjunction with other procedures to satisfy commonly encountered test constraints. In regard to a test blueprint, the default is to incorporate no constraints. However, three procedures are available to ensure that a test blueprint is met. Test items can either be chosen at random from different content areas to satisfy pre-specified limits, or from the multidimensional priority index approach described by Cheng and Chang (2009). To control item exposure rate, SimuMCAT supports the Symptom and Hetter (1985) procedure as well as a simpler approach that puts a maximum usage limit on the exposure rate of all items. Finally, for MCAT simulations in which the test length varies per examinees, two different stopping rules are supported: one based on achieving pre-specified standard errors of measurement; the other based on the predicted reduction in standard error (PSER; Yao, 2013).

Two methods of estimating MIRT abilities are supported in SimuMCAT. When items are selected using the Bayesian version of the selection procedure, maximum a posteriori (MAP) estimates are produced; otherwise maximum likelihood (ML) estimates are computed. Yao (2010) found that MAP estimates work best when estimating domain abilities. To reduce the bias inherent in many Bayesian procedures, the user can use a more non-informative prior.

With all the comprehensive features provided, SimuMCAT remains a user-friendly software package. Users simply need to modify the .txt and .par input files to introduce different conditions. Once the software is run, output files are self-explanatory with detailed information that can be used to analyze latent ability score recovery and item type usage under each item selection method. Runtime for the simulations can vary based on the item selection method, with noticeably longer time when test reliability coefficients are computed. When exact replications are desired, a constant random seed in the .txt input file can be used. For any combination of item selection method, set of content constraints, test lengths, and other factors generally varied per simulation, there are two output files (.ss and .par) that can be conveniently analyzed by directly reading them into other statistical software or data processing package such as R or SAS. For example, to analyze bias, absolute bias, and correlation of domain ability estimates with the true generating values, an array can be created in R that contains ability estimates of all examinees in all dimensions and across replications. Summary statistics can then be produced for further analyses.

Conclusion

SimuMCAT is a comprehensive software package that supports MCAT by providing realistic test constraints that might be encountered in operational testing. The software's capability to produce not only ability estimates but also other relevant information test reliability coefficients greatly enhances its usefulness. In addition, the software extends many regular features of adaptive testing to include new methods such as item selection methods that are not based on Fisher or Kullback–Leibler information, as well as the priority index method for content constraint. This in turn allows for more comprehensive research studies to be conducted to understand the best practice for MCAT in different testing situations. Future research is needed to compare the performance of ability recovery when simulating adaptive tests using SimuMCAT and other software packages such as R's package MAT (Choi, 2013) or package catR (Magis & Raiche, 2011) in the unidimensional setting.

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

- Cheng, Y., & Chang, H.-H. (2009). The maximum priority index method for severely constrained item selection in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, 62, 369-383.
- Choi, S. W. (2013, August 29). Package "MAT." Retrieved from <http://cran.r-project.org/web/packages/MAT/MAT.pdf>
- Luecht, R. M. (1996). Multidimensional computerized adaptive testing in a certification or licensure context. *Applied Psychological Measurement*, 20, 389-404.
- Magis, D., & Raiche, G. (2011). catR: An R package for computerized adaptive testing. *Applied Psychological Measurement*, 35, 576-577.
- Mulder, J., & van der Linden, W. J. (2009). Multidimensional adaptive testing with optimal design criteria for item selection. *Psychometrika*, 74, 273-296.
- Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, 21, 25-36.
- Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61, 331-354.
- Sympton, J. B., & Hetter, R. D. (1985). Controlling item exposure rates in computerized adaptive testing. In *Proceedings of the 27th annual meeting of the Military Testing Association* (pp. 937-977), San Diego, CA: Navy Personnel Research and Development Center.
- Veldkamp, B. P., & van der Linden, W. J. (2002). Multidimensional adaptive testing with constraints on test content. *Psychometrika*, 67, 575-588.
- Wang, C., & Chang, H.-H. (2011). Item selection in multidimensional computerized adaptive testing—Gaining information from different angles. *Psychometrika*, 76, 363-384.
- Wang, C., Chang, H.-H., & Boughton, K. A. (2011). Kullback-Leibler information and its applications in multi-dimensional adaptive testing. *Psychometrika*, 76, 13-39.
- Wang, W. C., & Chen, P. H. (2004). Implementation and measurement efficiency of multidimensional computerized adaptive testing. *Applied Psychological Measurement*, 28, 295-316.
- Yao, L. (2003). BMIRT: Bayesian Multivariate Item Response Theory [Computer software]. Monterey, CA: Defense Manpower Data Center.
- Yao, L. (2004). LinkMIRT: Linking of Multivariate Item Response Model [Computer software]. Monterey, CA: Defense Manpower Data Center.
- Yao, L. (2010). *Multidimensional ability estimation: Bayesian or non-Bayesian*. Unpublished manuscript.
- Yao, L. (2011). SimuMCAT: Simulation of Multidimensional Computer Adaptive Testing [Computer software]. Monterey, CA: Defense Manpower Data Center.
- Yao, L. (2012). Multidimensional CAT item selection methods for domain scores and composite scores: Theory and applications. *Psychometrika*, 77, 495-523.
- Yao, L. (2013). Comparing the performance of five multidimensional CAT selection procedures with different stopping rules. *Applied Psychological Measurement*, 37, 3-23.
- Yao, L. (2014, April). *Optimal item pool generation and the performance of multidimensional CAT*. Paper presented at the 2014 meeting of the National Council on Measurement in Education, Philadelphia, PA.
- Yao, L., Pommerich, M., & Segall, D. (2013, April). *Using multidimensional CAT to administer a short, yet precise screening test*. Paper presented at the 2013 meeting of the National Council on Measurement in Education, San Francisco, CA.
- Yao, L., & Schwarz, R. D. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. *Applied Psychological Measurement*, 30, 469-492.