Title
Computerized Adaptive Testing and Multistage Testing with R

Instructors
David Magis (david.magis@ulg.ac.be)
Department of Education, University of Liege, Belgium

Duanli Yan (dyan@ets.org)
Educational Testing Service, Princeton (NJ), USA

Alina A. von Davier (avondavier@ets.org)
Educational Testing Service, Princeton (NJ), USA

Length
1 day

Abstract
The goal of this workshop is to provide a practical (and brief) overview of the theory on computerized adaptive testing (CAT) and multistage testing (MST), and illustrate the methodologies and applications using R open source language and several data examples. The implementations rely on the R packages catR and mstR that have been already or are being developed and include some of the newest research algorithms developed by the authors.

This workshop will cover several topics: the basics of R, theoretical overview of CAT and MST, CAT and MST designs, assembly methodologies, catR and mstR packages, simulations and applications.

The intended audience for the workshop is undergraduate/graduate students, faculty, researchers, practitioners at testing institutions, and anyone in psychometrics, measurement, education, psychology and other fields who is interested in computerized adaptive and multistage testing, especially in practical implementations of simulation using R.

Summary
Computerized adaptive testing (CAT) has become a very popular method of administering questionnaires, collecting data and on-the-fly scoring (van der Linden & Glas, 2010; Wainer, 2015). It has been used in many large scale assessments over last decades and is currently an important field of research in psychometrics. Multistage testing (MST), on the other hand, got increased popularity in recent years (Yan, von Davier, & Lewis, 2014).

Both approaches rely on the notion of adaptive testing: items are administered sequentially and selected optimally according to the current responses to the administered items. In other words, the selection of the next items to administer depends on some current, ad-interim estimation of ability that is based on previously administered items. The conceptual difference between CAT and MST is that with CAT, items are selected one after each other (among a large pool of available items) and the ability of the test taker is estimated after the administration of each item. In MST, however, items are included in predefined modules and
the selection of the subsequent modules is based on the performance on the previously administered modules, not on the single items (Magis, Yan, & von Davier, in preparation).

Both methods have advantages and drawbacks with respect to each other and to linear testing. However, their practical usefulness relies mostly on accurate implementations of algorithms to perform test assembly, optimal item or module selection, (IRT) scoring, stopping rules and reporting.

In CAT, several commercial software (CATSim, Adaptest…) exist and some open-source solutions for simulation studies exist, most of them implemented in the R software, among others the packages catR (Magis & Barrada, in press; Magis & Raîche, 2012) and mirtCAT (Chalmers, 2015) and the R-based software Firestar (Choi, 2009). In MST, MSTGen (Han, 2013) exists. Very recently, the R package mstR was developed to provide a tool for simulations in the MST context, similarly to the catR package for CAT framework.

The purpose of this workshop is threefold: (a) to provide a brief overview of CAT and MST approaches and outline their specificities, advantages and drawbacks with respect to linear testing, as well as their technical challenges; (b) to present the R packages catR and mstR, their options and performances, in a simulation study-oriented perspective; (c) to run several examples of CAT and MST with both packages as illustrations.

The workshop will be a mix of theoretical and practical content. Demonstrations of catR and mstR will be used to illustrate the theoretical framework. Participants are encouraged to bring their laptops with R being pre-installed (and possibly also the R packages catR and mstR, though this can be fixed at the beginning of the workshop). Although R is available under Windows, Linux/ UNIX and MacOS platforms, demos will be run under Windows 7. Hands-out and R scripts will be made available for the participants.

References
Outline
Morning
Presentation of CAT and MST frameworks
Presentation of catR
Examples using catR

Afternoon:
Presentations of mstR
Examples using mstR
Wrap-up

Instructor bios
Dr. David Magis is Research Associate of the “Fonds de la Recherche Scientifique – FNRS” at the Department of Education, University of Liège, Belgium. His specialization is statistical methods in psychometrics, with special interest in item response theory, differential item functioning and computerized adaptive testing. His research interests include both theoretical and methodological development as well as open source implementation and dissemination with the statistical software R. He is currently associate editor of the British Journal of Mathematical and Statistical Psychology and published numerous research papers in various psychometric journals. He is the main developer and maintainer of the packages catR and mstR, among others.

Dr. Duanli Yan is a Manager of Data Analysis and Computational Research for Automated Scoring group in the Research and Development division at Educational Testing Service (ETS). She is also an Adjunct Professor at Rutgers University. She had been the statistical coordinator for the EXADEP™ test, and the TOEIC® Institutional programs, a Development Scientist for innovative research applications and a Psychometrician for several operational programs. She received many awards including 2011 ETS Presidential Award, 2013 NCME Brenda Loyd award, and 2015 IACAT Early Career Award. She is a co-editor for volume Computerized Multistage Testing: Theory and Applications. She is also a co-author for book Bayesian Networks in Educational Assessment. She has been an invited symposium organizer and presenter at many conferences such as those of the National Council of Measurement in Education (NCME), International Association for Computerized Adaptive Testing (IACAT), and International Psychometrics Society (IMPS).

Dr. Alina A. von Davier is a Senior Research Director at Educational Testing Service (ETS). She also is an Adjunct Professor at Fordham University. At ETS she leads the Computational Psychometrics Research Center, where is responsible for developing a team of experts and a psychometric research agenda in support of next generation of assessments. Computational psychometrics, which includes machine learning and data mining techniques, Bayesian inference methods, stochastic processes and psychometric models are the main set of tools employed in her current work. She also works with psychometric models applied to educational testing: test score equating methods, item response theory models, adaptive testing. She published several books and numerous papers in peer reviewed journals. During her tenure at ETS she also led the operational psychometric work for the international large-scale English assessments, such as TOEFL® and TOEIC®. She edited a volume on test equating, Statistical Models for Test Equating, Scaling, and Linking, which won 2013 AERA Division D Significant Contribution to Educational Measurement and Research Methodology award. She is a co-editor for volume Computerized Multistage Testing: Theory and Applications.