Book Review

Applied Psychological Measurement

2015, Vol. 39(3) 245–246

© The Author(s) 2014

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0146621614559744

apm.sagepub.com

Yan, D., von Davier, A. A., & Lewis, C. (2014). *Computerized Multistage Testing: Theory and Applications*. Boca Raton, FL: Chapman & Hall. 492 pp. \$ 90.00, ISBN-13: 978-1466505773

Reviewed by: Mariana Curi, *University of São Paulo, São Carlos, Brazil*

DOI: 10.1177/0146621614559744

Computer-based testing is spreading widely in education because of its advantages over traditional paper-and-pencil testing, such as faster score reporting, improved test security, ability to create innovative item formats, and widening access to big sectors of the population (van der Linden & Glas, 2010). Computerized testing has grown further since the advent of e-learning, large-scale assessments (e.g., Programme for International Student Assessment [PISA], Graduate Record Examination [GRE], and Test of English as a Foreign Language [TOEFL]), and the availability of modern psychometric methods. The strategies for building computerized tests can be divided into two classes: linear tests, in which all students receive the same items in the same order, and adaptive tests, in which the selection of the items is affected by the examinee's previous answers. In traditional computerized adaptive testing (CAT), the adaptation occurs after each answered item to generate more efficient proficiency estimates (Wainer, 1990). However, there are some practical problems with item-level adaptive tests. For example, CAT requires complex item selection algorithms to satisfy content constraints and to control item exposure rates, and it does not enable the examinees to review item answers.

The book presents multistage testing (MST), a useful methodology for the design and administration of adaptive tests that exhibits several practical advantages over traditional CAT. MST differs from traditional CAT in the degree of test adaptation: In CAT, individual items are selected for sequential administration, and in MST, sets of items (named “modules” or “testlets”) are the unity of selection (Zenisky, Hambleton, & Luecht, 2010). A multistage test is divided into at least two stages, composed by one or more modules that differ in difficulty. The number of consecutive stages, the number of modules, and the paths individuals can follow between stages define a MST design. This structure gives more flexibility to test developers, and it overcomes some CAT problems with little effect on test efficiency (Patsula & Hambleton, 1999; Zenisky et al., 2010).

This is the first comprehensive volume on MST, written by well-known psychometric researchers and practitioners. The theory and the implementation are carefully described, with respect to both classical test theory and item response theory approaches. In addition, extensive in-text citations provide a broad literature review and a range practical applications of MST tailored for different purposes, as well as interesting challenges and methodological issues that need to be addressed in future research. Therefore, this book provides a valuable guide to professionals in the testing industry, as well as to academics in the field of educational and psychological assessment. However, the content is structured such that the reader is assumed to be knowledgeable about the basic concepts of classical test theory and item response theory.

The book is divided into six parts. The first part, which is composed of five chapters, explains general concepts, design structures, and implementation decisions in MST, including the development and maintenance of an item pool. The first chapter summarizes the content in each of the remaining chapters of the book. Given this fact, and the fact that there are multiple

authors, the introductions for each of the remaining chapters are slightly repetitive. It might be beneficial to read Section 5.1 of Chapter 5 immediately after Chapter 1 because it explains the structure of an MST design in greater detail and introduces some of the mathematical formulations used in Chapters 2, 3, and 4. The second part, which consists of Chapters 6 to 9, presents current test assembly methods for MST, emphasizing on-the-fly assembly as a modern and feasible choice to improve test precision. In Chapter 7, there is an interesting presentation of the shadow-test approach as a two-stage item selection procedure as well as a discussion as to how any conceivable testing format can be viewed as a special case of this procedure. Chapters 10 to 15 form the third part of the book. In these chapters, algorithms for routing, scoring, and parameter estimation are proposed. These algorithms are based on item response theory or non-parametric models. MST is also considered from a classification perspective, using multidimensional and cognitive diagnostic models. Chapters 16 to 19, grouped into Part 4, consider concepts of test reliability, validity, fairness, and security. Chapters 20 to 26 are allocated to the fifth part. Six of them describe important considerations and decisions that must be made for the successful implementation of a MST in real-world educational assessments, from design to delivery. For example, Chapter 21 is dedicated to how the GRE test, perhaps the most famous application of MST, successfully implements this algorithm. The last chapter of this part introduces two open-source software programs that are available to run MST simulations. The book closes with Chapter 27, which discusses some historical background pertaining to the development of MST, as well as the most recent additional enhances to the MST algorithm, such as automated scoring and automated item generation.

This edited volume mirrors the process of developing and implementing an MST from beginning to end. It takes into consideration the decisions that need to be made at each step including the questions that need to be considered in the decision-making process, as well as the various methodologies that exist to implement the MST design chosen. The authors succeed in merging theoretical concepts with the operational and implementation aspects of MST, making the reading fruitful and motivating. The strength of the book is that it addresses the needs of practitioners in the testing industry, while still remaining theoretical enough to be of interest to graduate students and scientific researchers.

References

- Patsula, L. N., & Hambleton, R. K. (1999, April). *A comparative study of ability estimates from computer-adaptive testing and multi-stage testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Quebec, Canada.
- van der Linden, W. J., & Glas, C. A. W. (Eds.). (2010). *Elements of computerized adaptive testing*. New York, NY: Springer.
- Wainer, H. (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Erlbaum.
- Zenisky, A. L., Hambleton, R. K., & Luecht, R. M. (2010). Multistage testing: Issues, designs, and research. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 355-372). New York, NY: Springer.