Psychological Test and Assessment Modeling, Volume 54, 2012 (4), 363-365

Special topic: Current issues in Educational and Psychological Measurement: Design, calibration, and adaptive testing (Part 1) Guest Editorial

Andreas Frey[1] *& Ulf Kröhne*[2]

In the last decades, Educational and Psychological Measurement was a very active field of academic research. Numerous new methods and procedures were developed and many of them are now used on a regular basis and/or are implemented in statistical software packages. But, even though a solid state of knowledge has been established in many areas of Educational and Psychological Measurement, new demands and requirements are calling for new methodological answers and specific analysis procedures. Several of these demands and requirements stem from Large-Scale Assessments (LSAs). In LSAs, very large samples are examined; often under the objectives of deriving sound comparisons between quite different populations like countries and of drawing far-reaching inferences. The general objective of the Programme for International Student Assessment (PISA), for example, is to answer the rather general question of how well prepared students are to participate in society. The combination of examining very large samples, the desire for the comparison of rather different populations, and the aim to infer far-reaching interpretations creates a couple of demanding methodological challenges. Important methodological challenges that have not yet been answered sufficiently concern aspects of complex test designs used to distribute test items to participants, the handling of unwanted item context effects on both item parameter estimates and test performance, the calibration of data sets assessed with complex study designs, and the application of computerized adaptive testing (CAT) in order to meet specific diagnostic needs.

The special topic "Current issues in Educational and Psychological Measurement: Design, calibration, and adaptive testing" of Psychological Test and Assessment Modeling assembles a series of research papers addressing current issues in these areas. The gen-

[1] *Correspondence concerning this article should be addressed to:* Andreas Frey, PhD, Friedrich-Schiller-University Jena, Am Planetarium 4, 07737 Jena, Germany; email: andreas.frey@uni-jena.de

[2] German Institute for International Educational Research (DIPF), Frankfurt am Main, Germany

eral methodological approach used in all papers is the Item Response Theory (IRT). The special topic is spread over two issues of Psychological Test and Assessment Modeling. This issue is the first part and includes five papers.

With the first paper entitled "Principles and procedures of considering item sequence effects in the development of calibrated item pools: Conceptual analysis and empirical illustration" Yousfi and Böhme (2012) concentrate on item context effects due to the position and the sequence in which they are presented in test booklets. After introducing a taxonomy of booklet designs, different booklet designs are compared with regards to the bias and efficiency of item parameter estimates for CAT within two simulation studies.

The second paper entitled "On the importance of using balanced booklet designs in PISA" by Frey and Bernhardt (2012) focusses on the balanced booklet design used in PISA from the year 2003 on. The effects of a systematic distortion of the balanced booklet design structure on estimates for reading performance in different sub-populations are examined. Additionally, the question as to whether students with special characteristics are more prone to be advantaged or disadvantaged by a balanced booklet design compared to an unbalanced booklet design is analyzed.

The third paper entitled "A multilevel item response model for item position effects and individual persistence" from Hartig and Buchholz (2012) explicitly examines item position effects using student responses from different countries assessed in PISA 2006. In contrast to Yousfi and Böhme (2012), who compare different booklet designs with regards to item parameter estimates within simulation studies, Hartig and Buchholz investigate individual differences in item position effects and their relationship with student performance in science.

The fourth paper entitled "Capitalization on chance in variable-length classification tests employing the Sequential Probability Ratio Test" contributed by Patton, Cheng, Yuan, and Diao (2012) deals with the classification of individuals by using the sequential probability ratio test (SPRT) as termination criterion within CAT. With two simulation studies, the authors examine the effects of calibration error in item parameter estimates on test length and classification accuracy for computerized classification testing with the SPRT.

The fifth paper entitled "Biased (conditional) parameter estimation of a Rasch model calibrated item pool administered according to a branched testing design" by Kubinger, Steinfeld, Reif, and Yanagida (2012) addresses the problem of bias in item parameter estimates with the conditional maximum likelihood approach when a branched testing design is used. The relevance of the bias in item parameter estimates with regards to test scores measured with the Adaptive Intelligence Diagnosticum (Kubinger, 2009) is analyzed and discussed.

The second part of the special topic will follow in the next issue of Psychological Test and Assessment Modeling. Nevertheless, as guest editors, we would already like to thank the contributing authors of the first part for their elaborated and highly interesting articles which gave us the opportunity to learn lots of interesting new aspects of Educational and Psychological Measurement while assembling the special topic.

References

Frey, A., & Bernhardt, R. (2012). On the importance of using balanced booklet designs in PISA. *Psychological Test and Assessment Modeling*, *54*, 397-417.

Hartig. J., & Buchholz, J. (2012). A multilevel item response model for item position effects and individual persistence. *Psychological Test and Assessment Modeling*, *54*, 418-431.

Kubinger, K. D. (2009). *Adaptives Intelligenz Diagnostikum - Version 2.2 (AID 2) samt AID 2-Türkisch* [Adaptive Intelligence Diagnosticum, AID 2-Turkey included]. Göttingen: Beltz.

Kubinger, K. D., Steinfeld, J., Reif, M., & Yanagida, T. (2012). Biased (conditional) parameter estimation of a Rasch model calibrated item pool administered according to a branched testing design. *Psychological Test and Assessment Modeling*, *54*, 450-461.

Patton, J. M., Cheng, Y., Yuan, K. H., & Diao, Q. (2012). Capitalization on chance in variable-length classification tests employing the Sequential Probability Ratio Test. *Psychological Test and Assessment Modeling*, *54*, 432-449

Yousfi, S., & Böhme, H. (2012). Principles and procedures of considering sequence effects in the development of calibrated item pools: Conceptual analysis and empirical illustration. *Psychological Test and Assessment Modeling*, *54*, 366-396.