Psychological Test and Assessment Modeling, Volume 60, 2018 (3), 325-326

Guest Editorial
Special Topic: Advances in Educational Measurement Part II

Andreas Frey[1,2], Christoph König[3] & Christian Spoden[3]

The core of educational measurement lies in the representation of individual differences between and within persons or other educationally relevant entities by means of numbers. This includes theory building, empirical research, and application of study designs, instruments, data collection, statistical analysis, and the usage of the results obtained. Educational measurement is a very active field of research. The special topic "Advances in Educational Measurement" comprises nine recent research papers that focus on the following four aspects of educational measurement:

- Computerized Adaptive Testing and Multistage Testing
- Analysis of Large-Scale Assessment Data
- Multilevel Structural Equation Modeling
- Bayesian Modeling

Six of these papers were published in the first part of the special topic (Frey, König, & Spoden, 2018). The three papers in this second part focus on the aspects Computerized Adaptive Testing and Multistage Testing and Analysis of Large-Scale Assessment Data. The first paper by Fink, Born, Spoden, and Frey (2018) is devoted to the Computerized Adaptive Testing and Multistage Testing aspect. It presents a novel continuous calibration strategy for computerized adaptive testing (CAT). Until now, computerized adaptive tests have typically been applied in situations with large numbers of test-takers, where it is feasible to conduct a separate calibration study. In situations in which one of these two requirements is not met (e.g., recurring written university exams; testing in psychological or educational research), CAT is not used even though it would be advantageous. The authors address this problem by describing and examining the newly developed continuous calibration strategy. This strategy does not rely on an existing item pool (as existing online calibration methods do) and can be applied even with small samples as long as the test is repeated several times. The results of their Monte-Carlo simulation study show that the method yields accurate ability estimates even for sample sizes as small as $N = 50$. Thus, it

[1]*Correspondence concerning this article should be addressed to:* Andreas Frey, Institute of Educational Science, Department of Research Methods in Education, Friedrich Schiller University Jena, Am Planetarium 4, 07743 Jena, Germany. email: andreas.frey@uni-jena.de
[2]Centre for Educational Measurement (CEMO) at the University of Oslo, Norway
[3]Friedrich Schiller University Jena

is possible to calibrate items continuously across several test cycles without a separate calibration study and to obtain accurate ability estimates, which results in the test becoming better and better each cycle. This should make CAT very attractive for test developers in application areas commonly associated with small sample sizes and/or in which a separate calibration study is not possible.

The second paper by Yamamoto, Khorramdel, and Shin (2018) connects the Computerized Adaptive Testing and Multistage Testing aspect with the Analysis of Large-Scale Assessment Data aspect. Yamamoto et al. describe how multistage testing (MST) is implemented in the operational international large-scale assessment PIAAC (Programme for the International Assessment of Adult Competencies) and they evaluate the approach. The authors work at the Educational Testing Service, the organization leading the international PIAAC consortium, and can thus provide a detailed insight into the methods of this large-scale assessment. In the paper, they describe the MST approach used in PIAAC, examine the comparability of item parameter estimates across countries and observed item position effects, and compare the efficiency of the MST design with conventional nonadaptive testing methods that use item response theory (IRT) models. The results show only small item position effects, a high level of comparability of item parameter estimates across countries, and a higher test efficiency in the MST design compared to the conventional nonadaptive testing.

The third paper by Buchholz and Hartig (2018) is devoted to the aspect of Analysis of Large-Scale Assessment Data. It focuses on the potential effects of ignoring existing compensatory relations between ability dimensions when scaling response data from large-scale assessments. Until now, this issue has largely been neglected in applications of multidimensional IRT models. The authors conducted two Monte-Carlo simulation studies to evaluate the impact on test scores when a compensatory model is incorrectly applied to partially compensatory data. The findings show that biased ability parameter estimates can result, especially if the population distribution deviates from a multivariate normal distribution. Even though the findings imply that the effects are generally negligible when true abilities have a multivariate normal distribution, biased ability estimates may also emerge in this case for individuals with very low or very high abilities. The paper underlines the importance of fitting the correct psychometric model to the modeled content and concludes the special topic “Advances in Educational Measurement”.

References

Buchholz, J., & Hartig, J. (2018). The impact of ignoring the partially compensatory relation between ability dimensions on norm-referenced test scores. *Psychological Test and Assessment Modeling, 61*, 369-385.

Fink, A., Born, S., Spoden, C., & Frey, A. (2018). A continuous calibration strategy for computerized adaptive testing. *Psychological Test and Assessment Modeling*, *61*, 327-346.

Frey, A., König, C., & Spoden, C. (2018). Guest editorial special topic: Advances in educational measurement. *Psychological Test and Assessment Modeling, 60*, 141-144.

Yamamoto, K., Khorramdel, L., & Shin, H. J. (2018). Introducing multistage adaptive testing into international large-scale assessments designs using the example of PIAAC. *Psychological Test and Assessment Modeling*, *61*, 347-368.