Psychological Test and Assessment Modeling, Volume 58, 2016 (4), 593-595

Guest Editorial
Special Issue: Current Methodological Issues in Educational Large-Scale Assessments – Part I

Matthias Stadler[1], *Samuel Greiff*[2] *& Sabine Krolak-Schwerdt*[2]

Educational Large-Scale Assessments (LSAs), such as the Programme for International Student Assessment (PISA; OECD, 2015), the Programme for the International Assessment of Adult Competencies (PIAAC; Schleicher, 2008), or the Trends in International Mathematics and Science Study (TIMSS; Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009), are the objects of a growing and highly active area of research. Particular efforts are being made regarding the analysis and interpretation of corresponding results. On the one hand, LSAs provide an invaluable pool of rich data that allow for the application of complex methods to answer empirical research questions that cannot be addressed by smaller-scale studies. On the other hand, the adequate use and interpretation of these data pose unique methodological challenges. These may include issues as diverse as dealing with assessment instruments in different languages and their applicability across different cultures, establishing measurement invariance between these different assessment conditions, handling missing data, figuring out how to reduce the long computation times that are needed for complex analyses, or figuring out how to use process data in computer-based assessments.

The complex structure and size of international LSA databases often cause researchers to hesitate. For example, the 2012 cycle of the PISA assessment included data from 510,000 children from 65 economies. International LSA data therefore differ in many ways from more traditional data sets. For instance, LSA data (including international surveys) are usually not sampled at random, and students are typically not given every available test item (Martin, Mullis, & Kennedy, 2007; OECD, 2009). Moreover, there are particularly challenging organizational differences that must be handled adequately as they influence the data; examples are different testing modalities and language barriers (Butler & Stevens, 2001). For data analysis, specific approaches that are not part of

[1] *Correspondence concerning this article should be addressed to:* Matthias Stadler, PhD, Lehrstuhl für Methoden der empirischen Bildungsforschung, Universität Regensburg, Universitätsstraße 31, 93053 Regensburg, Germany; email: matthias.stadler@ur.de

[2] University of Luxembourg

many university curricula are required. As a consequence, the complex data structure of LSAs is often neglected by researchers, thus leading to the application of inadequate analyses and methods (Rutkowski, Gonzalez, Joncas, & von Davier, 2010).

Within this special issue, we aim to provide an overview of the vast array of methodological challenges that come with LSA data as well as the current state of the art in tackling them. Considering the diversity of challenges, we convinced various experts on different aspects of LSA to contribute their latest work to this special issue. However, because there were so many contributions, we needed to split the special issue into two parts in order to cover the whole range of methodological challenges posed by LSA.

The first part of the special issue consists of four papers highlighting the diversity of challenges while simultaneously introducing potential solutions. The authors, all established experts in the field of LSA, demonstrate exciting new ways of handling the transition to computer-based testing, maintaining maximum measurement precision, and dealing with missing data.

In the first paper, titled "The transition to computer-based testing in large-scale assessments: Investigating (partial) measurement invariance between modes," Sarah Bürger, Ulf Kröhne, and Frank Goldhammer illustrate how investigating (partial) measurement invariance between modes can facilitate the transition to computer-based testing in LSAs. The authors present a multiple-group IRT model approach for analyzing mode effects on the test and item levels. In addition, they review instances where partial measurement invariance is sufficient for combining item parameters into one metric. Finally, they present an extension of the modeling approach to explain mode effects by means of item properties.

The second paper is titled "Differentiated assessment of mathematical competence with multidimensional adaptive testing" and was contributed by Anna Mikolajetz and Andreas Frey. It addresses the important issue of reduced construct complexity due to time restrictions in LSAs. To deal with this problem, the authors demonstrate the effective use of multidimensional adaptive testing (MAT). Using the example of the German Educational Standards in Mathematics, which describes 11 subdimensions of mathematical competence, the paper shows how using MAT provides a way to measure a very complex construct with sufficient precision without increasing test length. This research closes the current gap between theoretical underpinnings, which describe highly distinguished subdomains of constructs and the time constraints on actual measures in LSAs.

In the third paper of this special issue, "Modeling test context effects in longitudinal achievement data: Examining position effects in the longitudinal German PISA 2012 Assessment," using the longitudinal extension of the PISA 2012 assessment in Germany, Gabriel Nagy, Oliver Lüdtke, and Olaf Köller investigate how test context effects affect scores in LSAs. The authors propose an extension of the two-dimensional one-parameter item response model, which includes the effects of booklets (i.e., test forms) on item clusters (i.e., item bundles) that are allowed to vary between assessment occasions and groups (school types). This additional consideration of context effects opens up the derivation of achievement scores in LSAs that can be compared across time more adequately.

The final paper in this first section of the special issue by Jonathan Weeks, Matthias von Davier, and Kentaro Yamamoto is titled "Using response time data to inform the coding of omitted responses." With empirical data from the PIAAC study, the authors examine the use of response time information collected in computer-based assessments to more correctly interpret the coding of missing responses. The authors aim to identify item-specific timing thresholds via several logistic regression models that predict the propensity of responding rather than produce a missing data point. With this procedure, missing data in LSAs can be handled on a far more detailed level than previously possible.

We hope that readers will enjoy this first part of our special issue and find it helpful for their own research. Moreover, we are particularly grateful to Klaus Kubinger, the Editor-in-Chief of the journal for hosting this special issue in Psychological Test and Assessment Modeling. In the second part of this special issue, five papers will cover IRT-based approaches to current methodological issues in LSA.

References

Butler, F. A., & Stevens, R. (2001). Standardized assessment of the content knowledge of English language learners K-12: Current trends and old dilemmas. *Language Testing, 18*(4), 409-427.

Martin, M., Mullis, I., & Kennedy, A. (Eds.). (2007). *PIRLS 2006 technical report.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks. International Association for the Evaluation of Educational Achievement.* Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

Organisation for Economic Co-operation and Development. (2009). *PISA 2006 technical report.* Paris: Author.

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data issues in secondary analysis and reporting. *Educational Researcher, 39*(2), 142-151.

Schleicher, A. (2008). PIAAC: A new strategy for assessing adult competencies. *International Review of Education, 54*(5-6), 627-650.