Special Topic: Methodological advances in psychological and educational testing

Guest-Editorial

Matthias von Davier[1]

This special issue of *Psychological Test and Assessment Modeling* provides a series of articles on recent developments in modern psychometric methodologies. The topics cover a range of topics including item fits in international assessments, response time modeling, treatment of missing data, and model-based procedures for standard setting in assessments.

Each topic would be worth a special issue in its own right. We are lucky to have experts from these diverse, important fields coming together to jointly publish this issue. However, since this special issue contains more articles than a single issue of Psychological test and Assessment Modeling can hold it was decided to split the special issue into two pieces. The first part contains four methodological articles on response time models, missing data, and scale score comparability, and a second issue with a group of papers on standard setting methods.

The first paper is Oliveri, M. E. & von Davier, M. (2011). Investigation of Model Fit and Score Scale Comparability in International Assessments. Psychological Test and Assessment Modeling, 53 (3) 315-333. The research presented in this paper examines the improvement of model-fit for international assessments. The analyses compare the Rasch model and the 2PL using standard linking through concurrent calibration with a modified 2PL concurrent linking approach that allows partially unique item parameters for the 2PL, while ensuring a strong link across countries. This methodology is presented in the context of the PISA (Programme for International Student Assessment) survey.

The second paper is van der Linden, W. (2011). Modeling Response Times with Latent Variables: Principles and Applications. Psychological Test and Assessment Modeling, 53 (3) 334-358. This paper discusses the introduction of a general model that integrates response times and item responses to improve measurement. The article reviews key principles for probabilistic modeling of response times and responses at the same time. It

---

[1] *Correspondence concerning this article should be addressed to:* Matthias von Davier, PhD, Educational Testing Service, Rosedale Rd, Princeton, NJ 08541, USA; email: mvondavier@ets.org

also introduces key components of models necessary to successfully integrate accuracy and latency information.

The third paper is Lee, Y. S. & Chen, H. (2011). A Review of Recent Response-Time Analyses in Educational Testing. Psychological Test and Assessment Modeling, 53 (3) 359-379. It surveys applications of response times since 2000. For the past decade, computer-based testing has been introduced more and more in general practice. The ubiquitous availability of technology in classrooms and testing centers has led to the collection of response latency. Many studies have been carried out to investigate whether response times are of value when predicting accuracy.

The fourth paper is Hohensinn, C. & Kubinger, K. (2011). On the impact of missing values on item fit and the model validness of the Rasch model. Psychological Test and Assessment Modeling, 53 (3) 380-393, and it leads us to a different way of examining the available data. Along with looking at collateral variables such as response times, it is important to examine why responses have not been given at all. Non-response in surveys or high-stakes tests is inevitable. But most procedures to treat non-responses are somewhat ad hoc and do not take into account that students with different levels of ability may have varying tendencies when it comes to omitting responses. This paper looks at the effects of ignoring non-responses, treating non-responses as incorrect responses on model fits, and item fit when using the Rasch model.

The remaining articles form a somewhat aligned group addressing different aspects of standard-setting. The first paper provides the conceptual approach to complex assessments; the other two introduce model-based approaches to standards setting. These three papers will follow in the next issue 53 (4) of Psychological Test and Assessment Modeling.

The fifth paper, "Standard-Setting and Complex Performance Assessments: An Approach Aligned with Cognitive Diagnostic Models," presents an approach referred to as the cognitive analytical approach to standards setting. It is based on the rational framework of assessment construction plant evidence in the design and integrates it with the recent emergence of models for cognitive diagnosis. These models enable us to assess multiple skills in a single assessment and to estimate mastery of the skills based on test taker's responses.

The sixth paper, "Investigating the Saltus Model as a Tool for Setting Standards," takes the Saltus model as the basis for setting standards. This model seems like a natural candidate for this type of task because it is based on a stage-wise development paradigm. Standards could be naturally aligned with all stages as long as items are intended to measure the standards function in a way that is consistent with the predicted difficulty of the stages involved.

The seventh paper is titled, "Exploring Levels of Performance Using the Mixture Rasch Model for Standard Setting." While the previous paper used a constraint version of the mixture Rasch model, this paper uses the general mixture Rasch model without constraints. This paper shows simulation results to present the potential of the mixture Rasch model for use as a tool to set performance cut-points and thus aid us in the process of developing standards based on data-driven methodologies. In addition to identifying cut-

points, this methodology also allows the researcher to identify respondents that are borderline, that is, respondents for whom it cannot be determined with certainty whether they pass a certain cut-point.

All papers in this collection have been written with actual use for large-scale data analysis in mind. The methodologies presented here have been evaluated in simulations as well as through real data analysis. While it is true that not all approaches are appropriate for all applications, they do present interesting venues to explore. Researchers interested in improving their methodological toolkit may want to consider using one or more approaches presented in this issue.

I hope readers feel they learned something new as much I did when assembling this special issue.