

## Preface to the Special Issue

Selda Yıldırım<sup>1</sup>  
Abant İzzet Baysal University

Large-scale assessments, such as Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA), provide big data in the educational context. A researcher who wants to conduct a secondary analysis using this big data has to notice that analyses of this type of data require considering some technical complexities. Among these complexities are incomplete test designs, sample weights, stratified sampling, and conducting the analyses and calculating the standard errors with respect to this sampling methodology. Information on these technical details is provided in the international reports of the corresponding assessments, and also in the related literature (e.g., Olson, Martin, & Mullis, 2008; OECD, 2005; OECD, 2009; Rutkowski, Gonzalez, Joncas, & von Davier, 2010). The aim of this special issue of *Education and Science* is not to provide definitive explanation of these technical complexities; rather, it is to constitute an educational resource which brings together studies that consider this complexity in their own right. In addition, this special issue also includes papers on methodologies that can be used in large-scale assessments. After almost a ten-month of intensive work we are glad that we can share this special issue consisting of 12 papers.

It is possible to divide the papers in this issue roughly into two groups as the papers focusing on the methodology, and the papers focusing on the investigation of the relationships between variables in the context of large-scale assessments. Methodology-focused papers begin with Verhelst's [1] study which discusses the basics of Classical Test Theory (CTT) and Item Response Theory (IRT). Besides, in his study Verhelst points out some widespread misunderstandings in testing. The following three papers mainly examine whether the different language versions of the same tests administered in the corresponding countries function equivalently across these countries. In their study, Lyons-Thomas, Sandilands and Ercikan [2] examine the gender differential item functioning (DIF) across four countries. We consider this study as a good example of using DIF procedures as a tool to reveal some significant factors associated with test performance of examinees. Arim and Ercikan [3], among other significant points, draw also attention to the fact that when directions of DIF items in a test are evenly distributed between countries, the overall effect of DIF on the test score may not be observed. Finally, the following study by Yıldırım, Yıldırım and Verhelts [4], presents a recent perspective which might also provide a promising answer to the difficulty of detecting the overall effect of DIF at a certain group level.

Finally, the last two papers of this group present empirical results on two diverse practices that can be used in large-scale assessments. Yıldırım [5] presents findings using two IRT-based vertical scaling perspectives in various contexts, which is required to obtain comparable test results from different assessments. The last methodology-focused paper by Bilen, Hotaman, Ařkın and Büyüklü [6] demonstrates an application of a data-mining procedure to discover some hidden structures in a big data, which is particularly appropriate when a little or none is known in the related area.

---

<sup>1</sup> Editor of the Special Issue; Assist. Prof. Dr., Abant İzzet Baysal University, Faculty of Education, Gökkyöy, BOLU, Turkey; cet\_s@ibu.edu.tr

However, when there is literature consisting of a growing body of research or some established theorizing about relationships between variables, researchers should consider this knowledge in designing their own study. In this context, if a rationale is to be specified for clustering the last six papers in a group, this would be that all these papers focus on relationships between variables considering the related literature. The five of them use data from the TIMSS 2011. Kahraman [7], Yetişir [8], Atar [9] and Ceylan [10] examine the relationships between student and/or teacher characteristics and the TIMSS science achievement. Akyüz [11] investigates the student and school characteristics that might be related to TIMSS mathematics achievement. While these studies generally use the plausible values as an indicator of student achievement, Ceylan [10] from a different perspective uses the item difficulties. Finally, in her study Güvendir [12] examines how student and school characteristics are related to student achievement in Turkish language in Student Achievement Determination Exam (ÖBBS), which is a national assessment.

I hope this special issue will serve as a valuable reference to the researchers and policymakers. I particularly thank Ziya Selçuk for his support and encouragement throughout the whole process. I also extend my thanks to the editors of the *Education and Science* who contributed to this issue with their guidance. Finally I also thank all of the authors who submitted their work for review, and the reviewers for their hard work in this process of bringing this issue into existence.

#### References

- Olson, J., Martin, M., & Mullis, I. (Eds.). (2008). TIMSS 2007 technical report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Organisation for Economic Co-operation and Development. (2005). *PISA 2003 data analysis manual*. Paris: Author.
- 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.