IERI Monograph Series

Issues and Methodologies in Large-Scale Assessments

VOLUME 4

October 2011
IERI Monograph Series
Issues and Methodologies in Large-Scale Assessments

Volume 4 2011

TABLE OF CONTENTS

Introduction 5
Matthias von Davier and Dirk Hastedt

Age distribution and reading achievement configurations among fourth-grade students in PIRLS 2006 9
Michael O. Martin, Ina V. S. Mullis, and Pierre Foy

The influences of home language, gender, and social class on mathematics literacy in France, Germany, Hong, Kong, and the United States 35
Aminah Perkins, Laura Quaynor, and George Engelhard, Jr.

Hierarchical factor item response theory models for PIRLS: capturing clustering effects at multiple levels 59
Frank Rijmen

Diagnostic cluster analysis of mathematics skills 75
Yoon Soo Park and Young-Sun Lee

TEDS-M: Diagnosing teacher knowledge by applying multidimensional item response theory and multiple-group models 109
Sigrid Blömeke, Richard T. Houang, and Ute Suhl

PISA test format assessment and the local independence assumption 131
Christian Monseur, Ariane Baye, Dominique Lafontaine, and Valérie Quittre

Information for contributors 157
Editors
Matthias von Davier  Educational Testing Service
Dirk Hastedt  IEA Data Processing and Research Center

Associate Editor
Oliver Neuschmidt  IEA Data Processing and Research Center

Production Editor
Marta Kostek  IEA Data Processing and Research Center

Assistant Production Editor
Katrin Jaschinski  IEA Data Processing and Research Center

Reviewers
Daniel Bolt  University of Wisconsin-Madison
Trevor Bond  James Cook University
Roel J. Bosker  University of Groningen
Wilfried Bos  Technische Universität Dortmund
Li Cai  University of California
Chia-Yi Chiu  Rutgers, the State University of New Jersey
James E. Corter  Columbia University
Paul de Boeck  Universiteit van Amsterdam
Christine DeMars  James Madison University
Jeffrey A. Douglas  University of Illinois
Pierre Foy  Boston College
Eugenio Gonzalez  Educational Testing Service
Shelby Haberman  Educational Testing Service
Thomas Martens  Deutsches Institut für Internationale Pädagogische Forschung
Jessica Lynn Mislevy  University of Maryland
Christian Monseur  University of Liège
Monika Rosén  Göteborg University
David Rutkowski  Indiana University
J. K. (Jeroen) Vermunt  Tilburg University
Xueli Xu  Educational Testing Service
Introduction

Matthias von Davier (Editor)
Educational Testing Service

Dirk Hastedt (Editor)
IEA Data Processing and Research Center

For more than four years, the International Association for the Evaluation of Educational Achievement (IEA) and Educational Testing Service (ETS) have worked together in the IEA-ETS Research Institute (IERI). IERI undertakes activities focused on three broad areas of work: research studies related to the development and implementation of large-scale assessments of educational outcomes, professional development and training, and dissemination of research findings and information gathered through large-scale assessments. Part of IEA and ETS’s collaborative work focuses on improving the science of large-scale assessments.

In this fourth volume of our periodical, we have collected a number of what we consider are very interesting papers that contribute to discussion about the analysis and findings of such assessments. Each paper has in common its authors’ efforts to deal with very specific and applied problems of analyzing international large-scale assessment data. Two of the papers use the IEA Progress in Reading Literacy Study (PIRLS) data, one uses the IEA Trends in International Mathematics and Science Study (TIMSS) data, two more employ OECD Programme for International Student Achievement (PISA) data, and the remaining paper employs IEA Teacher Education and Development Study in Mathematics (TEDS-M) data. All six papers contribute to discussions on methods used to analyze large-scale assessment data by critically reviewing the designs and analyses of these surveys and their data.

In the first paper, Age Distribution and Reading Achievement Configurations among Fourth-Grade Students in PIRLS 2006, authors Michael Martin, Ina Mullis, and Pierre Foy use PIRLS 2006 reading achievement results for fourth-grade students in 36 countries for their analyses. They show how variations in countries’ policies on age of school entry and on promotion and retention make it difficult to determine a straightforward relationship between age within grade and achievement across countries. Although, in some countries, older fourth-grade students have higher
achievement than younger students, older students do not necessarily perform better. The paper presents an interesting application of a regression discontinuity technique that the authors performed on data from countries with a strict cut-off for student progression from grade to grade.

The second paper, *The Influences of Home Language, Gender, and Social Class on Mathematics Literacy in France, Germany, Hong Kong, and the United States*, authors Aminah Perkins, Laura Quaynor, and George Engelhard, Jr. used data from four countries in order to analyze the relationships between home language, gender, and social class, and mathematics literacy. The data used came from PISA 2003. The authors used Rasch measurement and applied a clever approach wherein item difficulties were fixed to the operational valued from PISA. They also paid special attention to person response functions within the context of their analyses. The authors concluded that there are important differences between the groups under study. They suggested that their analyses should be replicated with data from more recent PISA cycles, and that this work should employ more recent estimation techniques and involve a larger group of participating countries.

In the third paper, *Hierarchical Factor Item Response Theory Models for PIRLS: Capturing Clustering Effects at Multiple Levels*, author Frank Rijmen presents a sophisticated analysis of data from an international large-scale assessment. He shows how modern psychometric methodologies can be applied to even large data-sets with a complex linking structure across test forms as well as complex dependencies between items due to the passage structure of the reading assessment. The use of forward-looking methodologies, such as the one that Rijmen presents, is becoming ever more important in terms of aiding our understanding of increasingly complex assessments that employ technology and more authentic item formats.

*Diagnostic Cluster Analysis of Mathematics Skills*, the fourth paper, is written by Yoon Soo Park and Young-Sun Lee and draws on TIMSS 2007 fourth-grade mathematics data. The authors present an alternative means of gaining a better understanding of how different skill domains are related across countries. The authors do not apply clustering to achieve classification of individuals. Instead, they classify skills into more homogeneous groups with respect to country selection. This process requires assigning each item the number of skills that are required to solve the item. When doing this, the authors drew on substantive support from content experts, such that they were able to provide the clustering solution for the required skills into more homogeneous clusters of skills for a number of countries.

The fifth paper, *TEDS-M: Diagnosing Teacher Knowledge by Applying Multi-dimensional Item Response Theory and Multiple-Group Models*, by Sigrid Blömeke, Richard Houang, and Ute Suhl, presents an analysis of TEDS-M data. The authors conducted this analysis in order to compare different multidimensional Rasch-type IRT models and to consider their utility in establishing latent variables suitable for representing teachers’ pedagogical content knowledge. The authors also address the question of population invariance and how the multidimensional models applied can help to inform research with respect to this issue.
PISA Test Format Assessment and the Local Independence Assumption, the sixth paper, by Christian Monseur, Ariane Baye, Dominique Lafontaine, and Valérie Quittre, provides an interesting application of measures that could be called, following Molenaar (1983), fit diagnostics. The authors explore whether local dependency is an issue in international large-scale assessments and whether taking the dependencies into account may allow us to better understand the results of these assessments.

We hope you will find reading these papers as interesting as we did. We also hope that you will consider supporting this periodical by submitting your own methodological research on international large-scale assessments to IERI.

Reference

About IEA
The International Association for the Evaluation of Educational Achievement (IEA) is an independent, non-profit, international cooperative of national research institutions and governmental research agencies. Through its comparative research and assessment projects, IEA aims to:

• Provide international benchmarks that can assist policymakers to identify the comparative strengths and weaknesses of their education systems;
• Provide high-quality data that will increase policymakers’ understanding of key school-based and non-school-based factors that influence teaching and learning;
• Provide high-quality data that will serve as a resource for identifying areas of concern and action, and for preparing and evaluating educational reforms;
• Develop and improve the capacity of educational systems to engage in national strategies for educational monitoring and improvement; and
• Contribute to development of the worldwide community of researchers in educational evaluation.

Additional information about IEA is available at www.iea.nl and www.iea-dpc.de.

About ETS
ETS is a non-profit institution whose mission is to advance quality and equity in education by providing fair and valid assessments, research, and related services for all people worldwide. In serving individuals, educational institutions, and government agencies around the world, ETS customizes solutions to meet the need for teacher professional development products and services, classroom and end-of-course assessments, and research-based teaching and learning tools. Founded in 1947, ETS today develops, administers, and scores more than 24 million tests annually in more than 180 countries, at over 9,000 locations worldwide.

Additional information about ETS is available at www.ets.org.