IERI Monograph Series

Issues and Methodologies in Large-Scale Assessments

VOLUME 2

October 2009

A joint publication between the International Association for the Evaluation of Educational Achievement (IEA) and Educational Testing Service (ETS)

Copyright © 2009 IEA-ETS Research Institute

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, electrostatic, magnetic tape, mechanical, photocopying, recording or otherwise without permission in writing from the copyright holder.

ETS and the ETS logo are registered trademarks of Educational Testing Service (ETS) in the USA and other countries worldwide and used under license by the IEA-ETS Research Institute (IERI). IEA and the IEA logo are trademarks and registered trademarks of the International Association for the Evaluation of Educational Achievement and used under license by the IEA-ETS Research Institute (IERI).

ISBN 978-0-88685-404-1

Copies of this publication can be obtained from:

IERInstitute
IEA Data Processing and Research Center
Mexikoring 37
22297 Hamburg,
Germany

IERInstitute
Educational Testing Service
Mail Stop 02-R
Princeton, NJ 08541,
United States

By email: ierinstitute@iea-dpc.de
Free downloads: www.ierinstitute.org

Copyedited by Paula Wagemaker Editorial Services, Christchurch, New Zealand
Design and production by Becky Bliss Design and Production, Wellington, New Zealand

IERI Monograph Series
Issues and Methodologies in Large-Scale Assessments

Volume 2 **2009**

TABLE OF CONTENTS

Introduction 5
Matthias von Davier and Dirk Hastedt

What are plausible values and why are they useful? 9
Matthias von Davier, Eugenio Gonzalez, and Robert J. Mislevy

Teachers' qualifications and their impact on student achievement: Findings from TIMSS 2003 data for Israel 37
Ruth Zuzovsky

Equivalence of item difficulties across national versions of the PIRLS and PISA reading assessments 63
Aletta Grisay, Eugenio Gonzalez, and Christian Monseur

Curriculum coverage and scale correlation on TIMSS 2003 85
Juliane Hencke, Leslie Rutkowski, Oliver Neuschmidt, and Eugenio Gonzalez

Questionnaire construct validation in the International Civic and Citizenship Education Study 113
Wolfram Schulz

Cluster analysis for cognitive diagnosis: An application to the 2001 PIRLS reading assessment 137
Chia-Yi Chiu and Minhee Seo

Variance estimation for NAEP data using a comprehensive resampling-based approach: An application of cognitive diagnostic models 161
Chueh-an Hsieh, Xueli Xu, and Matthias von Davier

Information for contributors 175

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

Reviewers

Roel Bosker	Hoogleraar Onderwijskunde
Claus H. Carstensen	Universität Kiel
Linda Cook	Educational Testing Services
Kai Cortina	University of Michigan
Michael Dean	International Baccalaureate, New York Teachers College
Karen Draney	University of California, Berkeley
Barbara Drechsel	Universität Kiel
Timo Ehmke	Universität Kiel
Howard Everson	Fordham University
Brian Junker	Carnegie Mellon University
Ann Kennedy	Boston College
Roy Levy	Arizona State University
Michael O. Martin	Boston College
Ivailo Partchev	Universität Jena
Carla Rampichini	Università degli Studi di Firenze
Frank Rijmen	Educational Testing Service
Andre A. Rupp	University of Maryland
David Rutkowski	IEA Data Processing and Research Center
Wolfram Schulz	Australian Council for Educational Research
Knut Schwippert	Universität Hamburg
Sandip Sinharay	Educational Testing Service

Production Editor

Marta Kostek	IEA Data Processing and Research Center

Assistant Production Editor

Katrin Jaschinski	IEA Data Processing and Research Center

Introduction

Matthias von Davier (Editor)
Educational Testing Service

Dirk Hastedt (Editor)
IEA Data Processing and Research Center

We are pleased to present Volume 2 of the *IERI Monograph Series*.

In 2007, IEA and ETS decided to establish the IEA-ETS Research Institute (IERI) with a focus on improving the science of large-scale assessments. IERI undertakes activities around three broad areas of work that include research studies related to the development and implementation of large-scale assessments; professional development and training; and dissemination of research findings and information gathered through large-scale assessments. Since IERI's establishment, many activities have taken place in the pursuit of the institute's mission.

One such activity is the institute's biannual training academies, which typically see 20 to 25 researchers from around the world attending each session. These academies provide researchers with training on the use of international large-scale assessment databases, and on more advanced statistical techniques for analyzing these data. To date, a total of six academies have been held. By the time this volume reaches you, the seventh academy will have taken place.

In addition to the academies, IERI has sponsored training on the use of large-scale assessment databases at a number of international education research conferences, such as those conducted by the American Educational Research Association (AERA), the Comparative & International Education Society (CIES), the European Educational Research Association (EERA), the International Association for Educational Assessment (IAEA), the International Meeting of the Psychometric Society (IMPS), and the National Council for Measurement in Education (NCME). IERI staff have also conducted similar training seminars for a number of other organizations in a variety of countries.

This monograph series represents, in part, the outcome of the training and research activities undertaken and supported by IERI. This volume begins with an examination of the use and usefulness of plausible values. In this paper, Matthias von Davier,

Eugenio Gonzalez, and Robert J. Mislevy present the rationale for using plausible values when estimating ability distributions in large-scale assessments. They discuss examples drawn from existing large-scale assessments, as well as examples based on simulations. The authors also offer arguments in support of the use of plausible values and the use of plausible value methodology.

The research paper by Ruth Zuzovsky deals with the interrelations between teacher preparation and professional development as it relates to student achievement. In her paper, Zuzovsky validates some assumptions regarding the interrelationship among these variables through reference to data from the Trends in International Mathematics and Science Study 2003 (TIMSS 2003). Results from this paper help inform the current debate surrounding teacher preparation and professional development and its effect on student achievement.

The paper by Aletta Grisay, Eugenio Gonzalez, and Christian Monseur tackles a topic that is at the heart of the comparability of large-scale assessments. The paper examines the equivalence of item difficulties across national versions of international large-scale assessments. The authors conducted their analyses using data from the Programme for International Student Assessment 2000 (PISA 2000) and Progress in Reading Literacy 2001 (PIRLS 2001) reading surveys.

The paper by Juliane Hencke, Leslie Rutkowski, Oliver Neuschmidt, and Eugenio Gonzalez focuses on the relationship between curriculum coverage and the results of TIMSS 2003. The extent to which results are comparable across participating countries is also at the heart of questions related to comparability in international assessments. This paper provides evidence to help us better understand the effect of curriculum coverage as it relates to the stability and comparability of results from these cross-national assessments.

The paper by Wolfram Schulz presents us with a study on construct validation in the International Civics and Citizenship Education Study. The paper presents and describes some of the procedures used in this study to validate the scales used during the pilot administration, and provides us with some preliminary results. The authors discuss the extent to which classical item statistics, factor analysis, and item response modeling help us assess the construct validity of questionnaire data obtained from international studies.

The final two papers, "Cluster Analysis for Cognitive Diagnosis: An Application to the 2001 PIRLS Reading Assessment," by Chia-Yi Chiu and Minhee Seo, and "Variance Estimation for NAEP Data Using a Resampling-based Approach: An Application of Cognitive Diagnostic Models," by Chueh-an Hsieh, Xueli Xu, and Matthias von Davier close this issue. The two papers present state-of-the-art statistical procedures that provide answers to the demands imposed on large-scale assessments to report more diagnostic information about the groups of examinees, and their cognitive profiles.

While Chiu and Seo focus on the cognitive diagnostic methods themselves, Hsieh, Xu, and von Davier present, via use of a multidimensional model for item response data, an application of a jackknifing approach to variance estimation of ability inferences for group of students. The data utilized to demonstrate the approach came from the

National Assessment of Educational Progress (NAEP). In contrast to the operational approach used in NAEP, where plausible values are used to make ability inferences, the approach presented in this paper re-estimates all parameters of the model, and makes ability inferences based on replicate samples using jackknife estimates without using plausible values.

We hope you enjoy reading and learning from the papers in this volume. Comments may be sent directly to contributing authors, or to the IERI email address: ierinstitute@iea-dpc.de. Information on submitting papers can be found on the last two pages of this volume or online at http://www.ierinstitute.org.

ABOUT THE 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 may assist policymakers identify the comparative strengths and weaknesses of their education systems;
- Provide high-quality data that will increase policy-makers' 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.