

Methodology of Educational Measurement and Assessment

Randy E. Bennett  
Matthias von Davier *Editors*

# Advancing Human Assessment

The Methodological, Psychological and  
Policy Contributions of ETS



Springer Open

# **Methodology of Educational Measurement and Assessment**

## **Series editors**

Bernard Veldkamp, Research Center for Examinations and Certification (RCEC),  
University of Twente, Enschede, The Netherlands

Matthias von Davier, National Board of Medical Examiners (NBME), Philadelphia,  
USA<sup>1</sup>

---

<sup>1</sup>This work was conducted while M. von Davier was employed with Educational Testing Service.

This book series collates key contributions to a fast-developing field of education research. It is an international forum for theoretical and empirical studies exploring new and existing methods of collecting, analyzing, and reporting data from educational measurements and assessments. Covering a high-profile topic from multiple viewpoints, it aims to foster a broader understanding of fresh developments as innovative software tools and new concepts such as competency models and skills diagnosis continue to gain traction in educational institutions around the world. *Methodology of Educational Measurement and Assessment* offers readers reliable critical evaluations, reviews and comparisons of existing methodologies alongside authoritative analysis and commentary on new and emerging approaches. It will showcase empirical research on applications, examine issues such as reliability, validity, and comparability, and help keep readers up to speed on developments in statistical modeling approaches. The fully peer-reviewed publications in the series cover measurement and assessment at all levels of education and feature work by academics and education professionals from around the world. Providing an authoritative central clearing-house for research in a core sector in education, the series forms a major contribution to the international literature.

More information about this series at <http://www.springer.com/series/13206>

Randy E. Bennett • Matthias von Davier  
Editors

# Advancing Human Assessment

The Methodological, Psychological  
and Policy Contributions of ETS



Springer Open

*Editors*

Randy E. Bennett  
Educational Testing Service (ETS)  
Princeton, NJ, USA

Matthias von Davier  
National Board of Medical Examiners  
(NBME)  
Philadelphia, PA, USA

CBAL, E-RATER, ETS, GRADUATE RECORD EXAMINATIONS, GRE, PRAXIS, PRAXIS I, PRAXIS II, PRAXIS III, PRAXIS SERIES, SPEECHRATER, SUCCESSNAVIGATOR, TOEFL, TOEFL IBT, TOEFL JUNIOR, TOEIC, TSE, TWE, and WORKFORCE are registered trademarks of Educational Testing Service (ETS). C-RATER, M-RATER, PDQ PROFILE, and TPO are trademarks of ETS.

ADVANCED PLACEMENT PROGRAM, AP, CLEP, and SAT are registered trademarks of the College Board. PAA is a trademark of the College Board.

PSAT/NMSQT is a registered trademark of the College Board and the National Merit Scholarship Corporation.

All other trademarks are the property of their respective owners.



ISSN 2367-170X

ISSN 2367-1718 (electronic)

Methodology of Educational Measurement and Assessment

ISBN 978-3-319-58687-8

ISBN 978-3-319-58689-2 (eBook)

DOI 10.1007/978-3-319-58689-2

Library of Congress Control Number: 2017949698

© Educational Testing Service 2017. This book is an open access publication

**Open Access** This book is licensed under the terms of the Creative Commons Attribution-NonCommercial 2.5 International License (<http://creativecommons.org/licenses/by-nc/2.5/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

This work is subject to copyright. All commercial rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Foreword

Since its founding in 1947, Educational Testing Service (ETS), the world's largest private nonprofit testing organization, has conducted a significant and wide-ranging research program. The purpose of *Advancing Human Assessment: The Methodological, Psychological and Policy Contributions of ETS* is to review and synthesize in a single volume the extensive advances made in the fields of educational and psychological measurement, scientific psychology, and education policy and evaluation by researchers in the organization.

The individual chapters provide comprehensive reviews of work ETS researchers conducted to improve the science and practice of human assessment. Topics range from test fairness and validity to psychometric methodologies, statistics, and program evaluation. There are also reviews of ETS research in education policy, including the national and international assessment programs that contribute to policy formation. Finally, there are extensive treatments of research in cognitive, developmental, and personality psychology.

Many of the developments presented in these chapters have become de facto standards in human assessment, for example, item response theory (IRT), linking and equating, differential item functioning (DIF), and confirmatory factor analysis, as well as the design of large-scale group-score assessments and the associated analysis methodologies used in the National Assessment of Educational Progress (NAEP), the Programme for International Student Assessment (PISA), the Progress in International Reading Literacy Study (PIRLS), and the Trends in International Mathematics and Science Study (TIMSS).

The breadth and the depth of coverage the chapters provide are due to the fact that long-standing experts in the field, many of whom contributed to the developments described in the chapters, serve as lead chapter authors. These experts contribute insights that build upon decades of experience in research and in the use of best practices in educational measurement, evaluation, scientific psychology, and education policy.

The volume's editors, Randy E. Bennett and Matthias von Davier, are themselves distinguished ETS researchers.

Randy E. Bennett is the Norman O. Frederiksen chair in assessment innovation in the Research and Development Division at ETS. Since the 1980s, he has conducted research on integrating advances in the cognitive and learning sciences, measurement, and technology to create new forms of assessment intended to have positive impact on teaching and learning. For his work, he was given the ETS Senior Scientist Award in 1996, the ETS Career Achievement Award in 2005, and the Distinguished Alumni Award from Teachers College, Columbia University in 2016. He is the author of many publications including “Technology and Testing” (with Fritz Drasgow and Ric Luecht) in *Educational Measurement* (4th ed.). From 2007 to 2016, he led a long-term research and development activity at ETS called the Cognitively Based Assessment of, for, and as Learning (CBAL<sup>®</sup>) initiative, which created theory-based assessments designed to model good teaching and learning practice.

Matthias von Davier is a distinguished research scientist at the National Board of Medical Examiners in Philadelphia, PA. Until January 2017, he was a senior research director at ETS. He managed a group of researchers concerned with methodological questions arising in large-scale international comparative studies in education. He joined ETS in 2000 and received the ETS Research Scientist Award in 2006. He has served as the editor in chief of the *British Journal of Mathematical and Statistical Psychology* since 2013 and is one of the founding editors of the SpringerOpen journal *Large-Scale Assessments in Education*, which is sponsored by the International Association for the Evaluation of Educational Achievement (IEA) and ETS through the IEA-ETS Research Institute (IERI). His work at ETS involved the development of psychometric methodologies used in analyzing cognitive skills data and background data from large-scale educational surveys, such as the Organisation for Economic Co-operation and Development’s PIAAC and PISA, as well as IEA’s TIMSS and PIRLS. His work at ETS also included the development of extensions and of estimation methods for multidimensional models for item response data and the improvement of models and estimation methods for the analysis of data from large-scale educational survey assessments.

ETS is proud of the contributions its staff members have made to improving the science and practice of human assessment, and we are pleased to make syntheses of this work available in this volume.

Research and Development Division  
Educational Testing Service,  
Princeton, NJ, USA

Ida Lawrence

# Preface

An edited volume on the history of the scientific contributions to educational research, psychology, and psychometrics made by the staff members of a nonprofit organization, Educational Testing Service (ETS), in Princeton, NJ, begs the questions: Why this self-inspection and who might benefit from that?

The answer can be found in current developments in these fields. Many of the advances that have occurred in psychometrics can be traced to the almost 70 years of work that transpired at ETS, and this legacy is true also for select areas in statistics, the analysis of education policy, and psychology.

When looking at other publications, be they conference proceedings, textbooks, or comprehensive collections like the *Handbook of Item Response Theory* (van der Linden 2016), the *Handbook of Test Development* (Lane et al. 2015), or *Educational Measurement* (4th ed.; Brennan 2006), one finds that many of the chapters were contributed by current and former ETS staff members, interns, or visiting scholars.

We believe that this volume can do more than summarize past achievements or collect and systematize contributions. A volume that compiles the scientific and policy work done at ETS in the years since 1947 also shows the importance, and the long-term effects, of a unique organizational form—the nonprofit measurement organization—and how that form can contribute substantially to advancing scientific knowledge, the way students and adults are assessed, and how learning outcomes and education policy more generally are evaluated.

Given the volume's purpose, we expect that it will be most attractive to those concerned with teaching, advancing, and practicing the diverse fields covered herein. It thus should make for an invaluable reference for those interested in the genesis of important lines of study that began or were significantly advanced at ETS.

Contributors to this volume are current and former ETS researchers who were asked to review and synthesize work around the many important themes that the organization explored over its history, including how these contributions have affected ETS and the field beyond. Each author brings his or her own perspective

---

This work was conducted while M. von Davier was employed with Educational Testing Service.



and writing style to the challenge of weaving together what in many cases constitutes a prodigious body of work. Some of the resulting accounts are thematically organized, while other authors chose more chronologically oriented approaches. In some chapters, individuals loom large simply because they had such significant impact on the topic at hand, while in other accounts, there were so many contributors that substantive themes offered a more sensible organizational structure. As a result, the chapters follow no single template, but each tells its own story in its own way.

The book begins with the reprint of a 2005 report by Randy E. Bennett that reviews the history of ETS, the special role that scientific research has played, and what that history might mean for the future.

With that opening chapter as context, Part I centers on ETS's contributions to the analytic tools employed in educational measurement. Chapters 2 and 3 by Tim Moses cover the basic statistical tools used by psychometricians and analysts to assess the quality of test items and test scores. These chapters focus on how ETS researchers invented new tools (e.g., confirmatory factor analysis), refined other tools (e.g., reliability indices), and developed versions of basic statistical quantities to make them more useful for assessing the quality of test and item scores for low- and high-stakes tests (e.g., differential item functioning procedures).

Chapter 4 by Neil J. Dorans and Gautam Puhan summarizes the vast literature developed by ETS researchers on one of the most fundamental procedures used to assure comparability of test scores across multiple test forms. This chapter on score linking is written with an eye toward the main purpose of this procedure, namely, to ensure fairness.

James E. Carlson and Matthias von Davier describe in Chap. 5 another staple of educational and psychological measurement that had significant roots at ETS. Their chapter on item response theory (IRT) traverses decades of work, focusing on the many developments and extensions of the theory, rather than summarizing the even more numerous applications of it.

In Chap. 6, Henry Braun describes important contributions to research on statistics at ETS. To name a few, ETS research contributed significantly to the methodology and discourse with regard to missing data imputation procedures, statistical prediction, and various aspects of causal inference.

The closing chapter in this part, Chap. 7 by Neil J. Dorans, covers additional topics relevant to ensuring fair assessments. The issues dealt with here go beyond that addressed in Chap. 4 and include procedures to assess differential item functioning (DIF) and give an overview of the different approaches used at ETS.

The chapters in Part II center on ETS's contributions to education policy and program evaluation. Two chapters cover contributions that built upon the developments described in Chaps. 2, 3, 4, 5, 6, and 7. Chapter 8 by Albert E. Beaton and John L. Barone describes work on large-scale, group-score assessments of school-based student populations. The chapter deals mainly with the National Assessment of Educational Progress (NAEP), describing the methodological approach and related developments. The following chapter by Irwin Kirsch, Mary Louise Lennon, Kentaro Yamamoto, and Matthias von Davier focuses on methods and procedures developed over 30 years at ETS to ensure relevance, comparability, and interpret-

ability in large-scale, international assessments of adult literacy. It describes how methods developed for NAEP were extended to focus on new target populations and domains, as well as to link assessments over time, across countries and language versions, and between modes of delivery.

Chapter 10 discusses longitudinal studies and related methodological issues. This contribution by Donald A. Rock reviews the series of important longitudinal investigations undertaken by ETS, showing the need to carefully consider assumptions made in such studies and in the vertical linking of tests associated with them. The approaches taken offer solutions to methodological challenges associated with the measurement of growth and the interpretation of vertical scales.

Besides the book's opening chapter, the only other contribution not written for the current volume is Chap. 11, by Samuel Ball. The chapter was originally published by ETS in 1979. It describes the pinnacle of work at ETS on large program evaluation studies, including the classic investigation of the effects of Sesame Street. We include it because Ball directed much of that work and his account offers a unique, firsthand perspective.

Part II concludes with Chap. 12, a review by Richard J. Coley, Margaret E. Goertz, and Gita Z. Wilder of the extensive set of projects, primary and secondary analyses, and syntheses produced on education policy. Those endeavors have ranged widely, from school finance analyses to help build more equitable funding approaches to uncovering the complex of factors that contribute to achievement gaps.

Part III concerns ETS's contributions to research in psychology and consists of three chapters. ETS's work in personality, social, cognitive, and developmental psychology was extensive. Much of it, particularly in the early years, centered on theory development, as well as on the invention and improvement of assessment methodology. The first two chapters, Chap. 13 by Lawrence J. Stricker and Chap. 14 by Nathan Kogan, respectively, cover the very broad range of investigations conducted in cognitive, personality, and social psychology. Chap. 15 by Nathan Kogan, Lawrence J. Stricker, Michael Lewis, and Jeanne Brooks-Gunn documents the large and extremely productive research program around the social, cognitive, and psychological development of infants and children.

The final part concerns contributions to validity. Chapter 16 by Michael Kane and Brent Bridgeman reviews ETS staff members' seminal work on validity theory and practice, most notably that of Samuel Messick. Donald E. Powers' Chap. 17 delves into the historically contentious area of special test preparation, an activity that can threaten or enhance the validity of test scores. The part closes with Isaac I. Bejar's Chap. 18, a wide-ranging examination of constructed-response formats, with special attention to their validity implications.

We end the book with Chap. 19, a synthesis of the material covered.

This book would not have been possible without the vision of Henry Braun, who suggested the need for it; Ida Lawrence, who supported it; Lawrence J. Stricker, whose contributions of time, effort, advice, and thought were invaluable; and the authors, who gave of their time to document the accomplishments of their colleagues, past and present. Also invaluable was the help of Kim Fryer, whose editorial and managerial skills saw the project to a successful completion.

We hope that this collection will provide a review worthy of the developments that took place over the past seven decades at ETS.

Princeton, NJ, USA  
September 2017

Randy E. Bennett  
Matthias von Davier

## References

- Brennan, R. L. (Ed.). (2006). *Educational measurement* (4th ed.). Westport: Praeger.
- Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2015). *Handbook of test development* (2nd ed.). New York: Routledge.
- van der Linden, W. J. (Ed.). (2016). *Handbook of item response theory: Models, statistical tools, and applications* (Vols. 1–3). Boca Raton: Chapman & Hall/CRC.