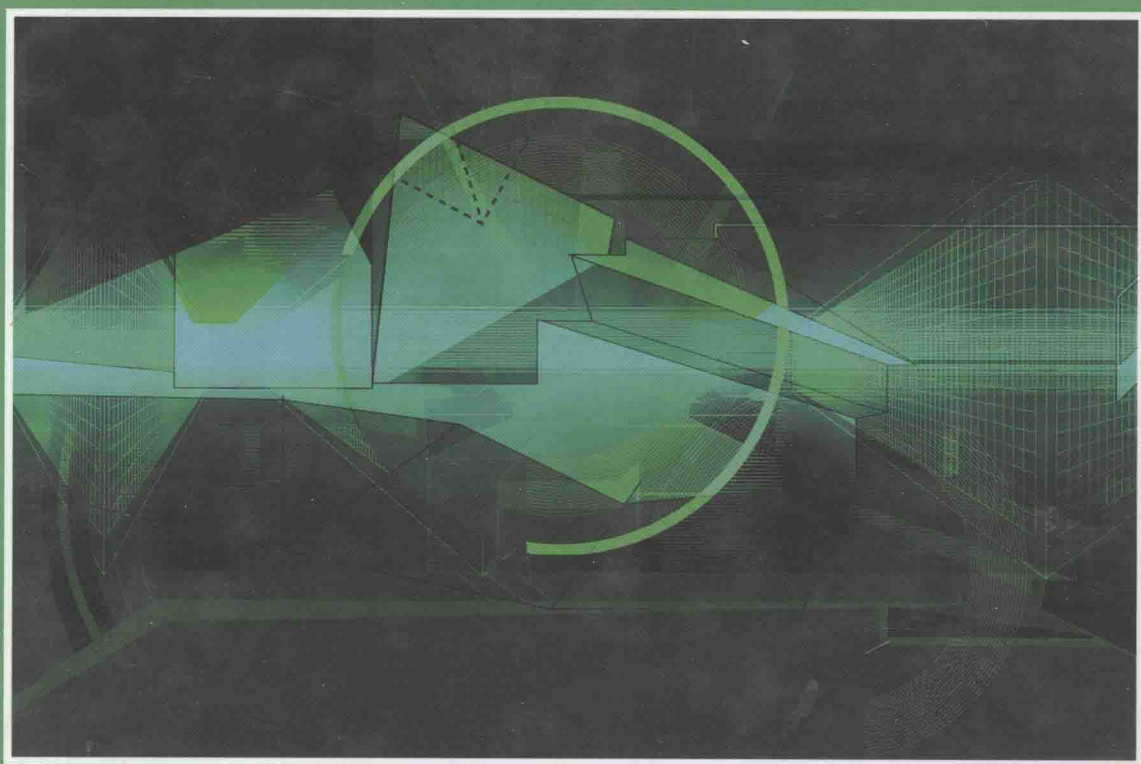


Measuring Psychological Constructs

Advances in Model-Based Approaches



Edited By

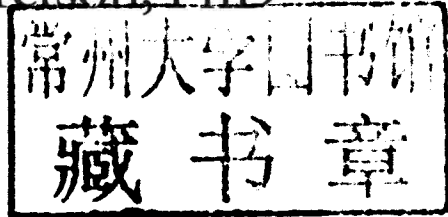
Susan E. Embretson, PhD

Measuring Psychological Constructs

Advances in Model-Based Approaches

Edited By

Susan E. Embretson, PhD



American Psychological Association

Washington, DC

Copyright © 2010 by the American Psychological Association. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, including, but not limited to, the process of scanning and digitization, or stored in a database or retrieval system, without the prior written permission of the publisher.

Published by
American Psychological Association
750 First Street, NE
Washington, DC 20002
www.apa.org

To order
APA Order Department
P.O. Box 92984
Washington, DC 20090-2984
Tel: (800) 374-2721;
Direct: (202) 336-5510
Fax: (202) 336-5502;
TDD/TTY: (202) 336-6123
Online: www.apa.org/books/
E-mail: order@apa.org

In the U.K., Europe, Africa, and the Middle East,
copies may be ordered from
American Psychological Association
3 Henrietta Street
Covent Garden, London
WC2E 8LU England

Typeset in New Century Schoolbook by Circle Graphics, Inc., Columbia, MD

Printer: Edwards Brothers, Ann Arbor, MI
Cover Designer: Minker Design, Bethesda, MD

The opinions and statements published are the responsibility of the authors, and such opinions and statements do not necessarily represent the policies of the American Psychological Association.

Library of Congress Cataloging-in-Publication Data

Measuring psychological constructs : advances in model-based approaches / Susan E. Embretson, editor. — 1st ed.

p. cm.

Based on presentations at a conference held in Feb. 2006 at the Georgia Institute of Technology in Atlanta, Georgia.

Includes bibliographical references and index.

ISBN-13: 978-1-4338-0691-9

ISBN-10: 1-4338-0691-6

1. Psychological tests. I. Embretson, Susan E.

BF176.M43 2010

150.28'7—dc22

2009022368

British Library Cataloguing-in-Publication Data

A CIP record is available from the British Library.

Printed in the United States of America
First Edition

Measuring Psychological Constructs

APA Science Volumes

Attribution and Social Interaction: The Legacy of Edward E. Jones
Best Methods for the Analysis of Change: Recent Advances, Unanswered Questions, Future Directions
Cardiovascular Reactivity to Psychological Stress and Disease
The Challenge in Mathematics and Science Education: Psychology's Response
Changing Employment Relations: Behavioral and Social Perspectives
Children Exposed to Marital Violence: Theory, Research, and Applied Issues
Cognition: Conceptual and Methodological Issues
Cognitive Bases of Musical Communication
Cognitive Dissonance: Progress on a Pivotal Theory in Social Psychology
Conceptualization and Measurement of Organism–Environment Interaction
Converging Operations in the Study of Visual Selective Attention
Creative Thought: An Investigation of Conceptual Structures and Processes
Developmental Psychoacoustics
Diversity in Work Teams: Research Paradigms for a Changing Workplace
Emotion and Culture: Empirical Studies of Mutual Influence
Emotion, Disclosure, and Health
Evolving Explanations of Development: Ecological Approaches to Organism–Environment Systems
Examining Lives in Context: Perspectives on the Ecology of Human Development
Global Prospects for Education: Development, Culture, and Schooling
Hostility, Coping, and Health
Measuring Patient Changes in Mood, Anxiety, and Personality Disorders: Toward a Core Battery
Occasion Setting: Associative Learning and Cognition in Animals
Organ Donation and Transplantation: Psychological and Behavioral Factors
Origins and Development of Schizophrenia: Advances in Experimental Psychopathology
The Perception of Structure
Perspectives on Socially Shared Cognition

Psychological Testing of Hispanics

Psychology of Women's Health: Progress and Challenges in Research and Application

Researching Community Psychology: Issues of Theory and Methods

The Rising Curve: Long-Term Gains in IQ and Related Measures

Sexism and Stereotypes in Modern Society: The Gender Science of Janet Taylor Spence

Sleep and Cognition

Sleep Onset: Normal and Abnormal Processes

Stereotype Accuracy: Toward Appreciating Group Differences

Stereotyped Movements: Brain and Behavior Relationships

Studying Lives Through Time: Personality and Development

The Suggestibility of Children's Recollections: Implications for Eyewitness Testimony

Taste, Experience, and Feeding: Development and Learning

Temperament: Individual Differences at the Interface of Biology and Behavior

Through the Looking Glass: Issues of Psychological Well-Being in Captive Nonhuman Primates

Uniting Psychology and Biology: Integrative Perspectives on Human Development

Viewing Psychology as a Whole: The Integrative Science of William N. Dember

APA Decade of Behavior Volumes

Acculturation: Advances in Theory, Measurement, and Applied Research

Aging and Cognition: Research Methodologies and Empirical Advances

Animal Research and Human Health: Advancing Human Welfare Through Behavioral Science

Behavior Genetics Principles: Perspectives in Development, Personality, and Psychopathology

Categorization Inside and Outside the Laboratory: Essays in Honor of Douglas L. Medin

Chaos and Its Influence on Children's Development: An Ecological Perspective

Child Development and Social Policy: Knowledge for Action

Children's Peer Relations: From Development to Intervention

Commemorating Brown: The Social Psychology of Racism and Discrimination

Computational Modeling of Behavior in Organizations: The Third Scientific Discipline
Couples Coping With Stress: Emerging Perspectives on Dyadic Coping
Developing Individuality in the Human Brain: A Tribute to Michael I. Posner
Emerging Adults in America: Coming of Age in the 21st Century
Experimental Cognitive Psychology and Its Applications
Family Psychology: Science-Based Interventions
Inhibition and Cognition
Measuring Psychological Constructs: Advances in Model-Based Approaches
Medical Illness and Positive Life Change: Can Crisis Lead to Personal Transformation?
Memory Consolidation: Essays in Honor of James L. McGaugh
Models of Intelligence: International Perspectives
The Nature of Remembering: Essays in Honor of Robert G. Crowder
New Methods for the Analysis of Change
On the Consequences of Meaning Selection: Perspectives on Resolving Lexical Ambiguity
Participatory Community Research: Theories and Methods in Action
Personality Psychology in the Workplace
Perspectivism in Social Psychology: The Yin and Yang of Scientific Progress
Primate Perspectives on Behavior and Cognition
Principles of Experimental Psychopathology: Essays in Honor of Brendan A. Maher
Psychosocial Interventions for Cancer
Racial Identity in Context: The Legacy of Kenneth B. Clark
The Social Psychology of Group Identity and Social Conflict: Theory, Application, and Practice
Strengthening Couple Relationships for Optimal Child Development: Lessons From Research and Intervention
Strengthening Research Methodology: Psychological Measurement and Evaluation
Transcending Self-Interest: Psychological Explorations of the Quiet Ego
Unraveling the Complexities of Social Life: A Festschrift in Honor of Robert B. Zajonc
Visual Perception: The Influence of H. W. Leibowitz

Contributors

- Lisbeth Åberg-Bengtsson, PhD**, Department of Education, University of Gothenburg, Gothenburg, Sweden
- William H. Batchelder, PhD**, School of Social Sciences, University of California, Irvine
- Isaac I. Bejar, PhD**, Educational Testing Service, Princeton, NJ
- Louis V. DiBello, PhD**, Learning Sciences Research Institute, University of Illinois, Chicago
- Susan E. Embretson, PhD**, Department of Psychology, Georgia Institute of Technology, Atlanta, GA
- Jan-Eric Gustafsson, PhD**, Department of Education, University of Gothenburg, Gothenburg, Sweden
- Robert A. Henson, PhD**, Educational Research Methodology Department, The University of North Carolina at Greensboro
- Eunice Jang, PhD**, Department of Curriculum, Teaching, and Learning, University of Toronto, Toronto, Ontario, Canada
- Rianne Janssen, PhD**, Department of Educational Sciences, University of Leuven, Leuven, Belgium
- George B. Macready, PhD**, College of Education, University of Maryland, College Park
- Steven P. Reise, PhD**, Department of Psychology, University of California, Los Angeles
- James S. Roberts, PhD**, School of Psychology, Georgia Institute of Technology, Atlanta
- Jürgen Rost, PhD**, IPN Institute for Science Education, University of Kiel, Kiel, Germany
- Louis A. Roussos, PhD**, Measured Progress, Dover, NH
- Lynne Steinberg, PhD**, Department of Psychology, University of Houston, Houston, TX
- Jonathan L. Templin, PhD**, Department of Educational Psychology and Instructional Technology, University of Georgia, Athens
- David Thissen, PhD**, L. L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill
- Matthias von Davier, PsyD**, Educational Testing Service, Princeton, NJ
- Niels G. Waller, PhD**, Department of Psychology, University of Minnesota, Minneapolis

Foreword

In early 1988, the American Psychological Association (APA) Science Directorate began its sponsorship of what would become an exceptionally successful activity in support of psychological science—the APA Scientific Conferences program. This program has showcased some of the most important topics in psychological science and has provided a forum for collaboration among many leading figures in the field.

The program has inspired a series of books that have presented cutting-edge work in all areas of psychology. At the turn of the millennium, the series was renamed the Decade of Behavior Series to help advance the goals of this important initiative. The Decade of Behavior is a major interdisciplinary campaign designed to promote the contributions of the behavioral and social sciences to our most important societal challenges in the decade leading up to 2010. Although a key goal has been to inform the public about these scientific contributions, other activities have been designed to encourage and further collaboration among scientists. Hence, the series that was the “APA Science Series” has continued as the “Decade of Behavior Series.” This represents one element in APA’s efforts to promote the Decade of Behavior initiative as one of its endorsing organizations. For additional information about the Decade of Behavior, please visit <http://www.decadeofbehavior.org>.

Over the course of the past years, the Science Conference and Decade of Behavior Series has allowed psychological scientists to share and explore cutting-edge findings in psychology. The APA Science Directorate looks forward to continuing this successful program and to sponsoring other conferences and books in the years ahead. This series has been so successful that we have chosen to extend it to include books that, although they do not arise from conferences, report with the same high quality of scholarship on the latest research.

We are pleased that this important contribution to the literature was supported in part by the Decade of Behavior program. Congratulations to the editors and contributors of this volume on their sterling effort.

Steven J. Breckler, PhD
Executive Director for Science

Virginia E. Holt
*Assistant Executive Director
for Science*

Preface

This book highlights several explanatory approaches to model-based measurement. These approaches not only extend rigorous psychometric methods to a variety of important psychological constructs but, more important, also have the potential to fundamentally change the nature of the constructs that are measured. The models in the exploratory approaches have special parameters or features that can represent important aspects of constructs that are not represented in standard item response theory models.

The chapters in this volume are based on presentations that were given at a conference, “New Directions in Measuring Psychological Constructs With Model-Based Approaches.” The conference was held in February 2006 at the Georgia Institute of Technology in Atlanta. The conference was jointly sponsored by the American Psychological Association and the Georgia Institute of Technology. The book features chapters by 11 internationally distinguished authors, who have varying perspectives on how measurement constructs are impacted by modern psychometric modeling approaches.

Special recognition must be given to individuals who contributed their time and effort to the conference and to the production of this volume. Dr. James Roberts assisted not only with managing the conference but also with reviewing chapters for this volume. Several graduate students in the Quantitative Psychology Program at the Georgia Institute of Technology also contributed to both the conference and the editing of the book. These students are Robert Daniel, Heather McIntyre, Hi Shin Shim, and Vanessa Thompson.

Finally, Marshall Picow, my husband, had a vital role in supporting me throughout the period of the conference and the preparation of the book. This volume probably would not have been possible without his unfailing and devoted support during my recovery from a serious accident that occurred during the week after the conference.

Measuring Psychological Constructs

Contents

Contributors	xi
Foreword	xiii
Preface	xv
1. Measuring Psychological Constructs With Model-Based Approaches: An Introduction	1
<i>Susan E. Embretson</i>	
Part I. Model-Based Approaches to Measuring Qualitative Differences Between Individuals	9
2. Mixture Distribution Item Response Theory, Latent Class Analysis, and Diagnostic Mixture Models	11
<i>Matthias von Davier</i>	
3. Skills Diagnosis for Education and Psychology With IRT-Based Parametric Latent Class Models	35
<i>Louis A. Roussos, Louis V. DiBello, Robert A. Henson, Eunice Jang, and Jonathan L. Templin</i>	
4. Cognitive Psychometrics: Using Multinomial Processing Tree Models as Measurement Tools	71
<i>William H. Batchelder</i>	
Part II. Model-Based Approaches to Isolating Entangled Constructs	95
5. Unidimensionality and Interpretability of Psychological Instruments	97
<i>Jan-Eric Gustafsson and Lisbeth Åberg-Bengtsson</i>	
6. Using Item Response Theory to Disentangle Constructs at Different Levels of Generality	123
<i>David Thissen and Lynne Steinberg</i>	

**Part III. Model-Based Approaches for Measuring Personality,
Psychopathology, and Attitudes From Self-Reports 145**

7. Measuring Psychopathology With Nonstandard
Item Response Theory Models: Fitting the
Four-Parameter Model to the Minnesota Multiphasic
Personality Inventory 147
Niels G. Waller and Steven P. Reise

8. MIXUM: An Unfolding Mixture Model to Explore the
Latitude of Acceptance Concept in Attitude Measurement 175
James S. Roberts, Jürgen Rost, and George B. Macready

**Part IV. Cognitive Psychometric Models for Interactive
Item Generation 199**

9. Recent Development and Prospects in Item Generation 201
Isaac I. Bejar

10. Modeling the Effect of Item Designs Within the
Rasch Model 227
Rianne Janssen

11. Cognitive Design Systems: A Structural Modeling
Approach Applied to Developing a Spatial Ability Test 247
Susan E. Embretson

Index 275

About the Editor 285

Measuring Psychological Constructs With Model-Based Approaches: An Introduction

Susan E. Embretson

More than a half century has passed since Cronbach (1957) made his well-known distinction between the correlational and the experimental disciplines in psychology. The disciplines were characterized not only by distinct methods but also by different kinds of constructs. Psychological measurement is almost entirely within the correlational discipline, which has individual differences as a primary focus. Certainly substantial changes have occurred within the two scientific disciplines and within psychological measurement over the past 50 years. But have the constructs that are measured using contemporary psychometric methods become better integrated with constructs that arise from the experimental discipline of psychology?

Within the measurement field over the past half century, item response theory (IRT) models and methods have replaced classical test theory (CTT) as the basis for developing many psychological and educational tests. IRT is model-based measurement, in that the individual item responses are modeled. Thus, IRT models include not only one or more estimates to represent the persons but also estimates to represent the psychometric properties of items. This contrasts sharply with CTT, in which the main target is total score, not item responses. Hence, item properties are not represented directly in the model.

The many practical, technical, and statistical advantages of IRT for developing psychological and educational tests have been published widely in both methodological and substantive journals. Further, IRT has been a major focus in numerous conferences and workshops, and these gatherings have resulted in many edited books. The properties of IRT models have also been given considerable attention. For example, *Rasch Models: Foundations, Recent Developments and Applications* (Fischer & Molenaar, 1995) especially highlights the formal quantitative aspects of the models. *Objective Measurement: Theory Into Practice, Volume 5* (Wilson & Engelhard, 2000) is an example of a series that contains a broad sampling of new applications and extensions of IRT models. Other books, such as *Computerized Adaptive Testing* (Wainer, 1990), focus specifically on a major practical advantage of IRT-based tests. Collectively, the many available edited volumes have made substantial contributions to psychometric methods and have further extended IRT into measurement in many substantive areas.

However, the integration of model-based approaches of IRT with substantive research on psychological constructs has received less attention. An edited book, *Test Validity* (Wainer & Braun, 1988), contains chapters that are focused on new approaches to explicating validity, which includes IRT model-based approaches. Although contemporary psychological theory is represented by a couple of articles in this book, the vast majority of articles concern methodological developments arising from within the correlational discipline. Another edited book that arose from a measurement-oriented conference focuses more directly on construct issues. A conference honoring Samuel Messick was held at Educational Testing Service in 1997. Although the papers from this conference (see Braun, Jackson, & Wiley, 2002) raised several issues about psychological constructs, they were not related to specific model-based measurement approaches. Furthermore, many papers at the conference could be characterized as primarily based on correlational methods of theory development.

The topics of two edited books are more directly relevant to interfacing measurement methods with contemporary psychological constructs. *Test Theory for a New Generation of Tests* (Frederiksen, Mislevy, & Bejar, 1993) introduced model-based measurement as providing new approaches to testing and new ways to interface with substantive theory. Few applications have yet been realized, even though the book preceded the present volume by more than 15 years. Unfortunately, such delays are not atypical in the slowly evolving world of psychological testing. Some important recent developments in model-based measurement are covered in *Explanatory Item Response Models* (De Boeck & Wilson, 2004), along with many interesting illustrative applications. A major goal of De Boeck and Wilson was to provide an integrated presentation of several recent IRT models that could be specified with common statistical framework (i.e., as non-linear mixed models). Hence, the scope of their book was necessarily limited.

Thus, despite many important developments in psychological measurement in the past half century, there is little evidence that the integration between measurement and the constructs that stem from psychological theory has progressed much since Cronbach (1957) noted the two separate disciplines of psychology. Perhaps the areas are fundamentally incompatible. This would be especially true if individual differences were never of interest in experimentally based theories. However, another possibility is that typical applications of IRT, like its predecessor CTT, may introduce constraints that have limited the applicability of psychometric methods to constructs of interest in contemporary psychology. That is, the most widely used psychometric methods are most appropriately applied when a single source of impact (i.e., underlying dimension) influences both persons and items.

The purpose of this volume, in part, is to present a broad spectrum of model-based measurement approaches that remove some of the constraints. Typical test development practices under both CTT and IRT require several assumptions that do not necessarily interface well with psychological constructs as conceptualized theoretically. That is, the test developer must assume that (a) the same construct can characterize responses of all persons, (b) items have identical psychometric properties when administered to different persons, (c) items are fixed entities with known stimulus content, (d) items are calibrated prior to test scoring, (e) item response probabilities are monotonically related to the trait to

be measured, and (f) internal consistency between items on a test indicates adequate assessment of a trait.

Although these assumptions seem fundamental to developing psychometrically rigorous tests, they have also functioned to define a narrow set of tasks and conditions that are deemed appropriate for measurement. One obvious consequence has been the popularity of multiple-choice tasks and related objective item formats because they more readily meet the constraints than do other tasks, such as constructed responses. Yet constructed response tasks and performance assessment are often regarded as more authentic measures that better represent conceptualizations of the domain or construct. The constraints also impact test content in other ways than item format. For example, empirical tryout of items involves evaluation and selection in terms of the constraints noted above. Item attrition is often quite high; a rate of 50% attrition is not atypical. With such high item attrition, the surviving items may not represent very well the original conceptualization of the task domain.

Several developments in model-based measurement have the potential to impact the nature of constructs that can be measured in psychology. Some of these developments remove one or more constraints, as described earlier, whereas other developments, such as explanatory psychometric models, permit a new level of integration of measurement and psychological theory. Many of these models (e.g., De Boeck & Wilson, 2004; Embretson, 1999; Glas & Van der Linden, 2003; Mislevy, Steinberg, & Almond, 2003; Roberts, Donoghue, & Laughlin, 2000; Rost, 1990, 1991; von Davier & Rost, 1995) have appeared in the psychometric literature but have not yet been available to the broader audience of psychologists.

The purpose of this volume is to highlight several explanatory approaches to model-based measurement that can impact the nature of the psychological constructs that can be measured with rigorous psychometric methods. In this book, model-based measurement approaches that are more appropriate for the constructs of interest in many substantive areas of psychology are explicated and illustrated with applications. New developments of model-based measurement in four different areas are included as follows: (a) model-based approaches to measuring qualitative differences between individuals; (b) model-based approaches to isolating entangled constructs; (c) model-based approaches for measuring personality, psychopathology, and attitudes from self-reports; and (d) cognitive psychometric models for interactive item generation during testing.

In Part I, *Model-Based Approaches to Measuring Qualitative Differences Between Individuals*, several different types of explanatory models are represented. Qualitative differences between persons in the nature of a construct may occur either in cognitive measurement or in personality and attitude measurement. A well-studied source of qualitative differences between groups is differential item functioning (DIF), which is a violation of the traditional psychometric constraint that items have identical properties for all examinees. A common psychometric procedure is to eliminate items showing DIF, which consequently narrows measurement of the construct. Another approach is to use model-based approaches with DIF items included but with group-specific parameters. Although the item domain is not narrowed by using this approach, it is nonetheless controversial because an individual's estimated score will be impacted by demographic variables, such as gender or ethnicity.

In contrast, a model-based solution has broad potential to identify classes of individuals who differ qualitatively on the construct. These classes are based not on demographics but on the pattern of item responses. Mixture IRT models (e.g., Rost, 1991; von Davier & Rost, 1995) can be applied when the test scores do not represent the same construct for different examinees. Such a test probably would have poor fit to a traditional IRT model because the assumption of unidimensionality is violated. The mixture IRT models identify latent classes of examinees for whom the construct differs qualitatively. Thus, both the score levels and the latent class memberships of the persons can provide important information about individual differences. To give an example, success in solving spatial tasks does not necessarily involve spatial analogue processing. Some spatial tasks may be solved by either analogue or verbal-analytic processes, which can have implications for the external correlates of test scores (Embretson, 2007). Of course, tasks that can be solved by more than one method could be eliminated from spatial ability measurement; however, the remaining tasks may be sufficiently restricted so as to adversely impact the theoretical scope of the construct. In chapter 2, Matthias von Davier presents an overview of mixture distribution IRT models. His chapter shows a progression of models, from unidimensional IRT models and latent class models to the mixture distribution IRT models, to handle varying assumptions about the nature of the construct.

Items that measure the same dominant trait can also differ qualitatively. Diagnostic IRT-based models, such as the fusion model (Roussos et al., 2007), can be used to relate qualitative features of items to item solving. In mathematics achievement, for example, qualitative information about the specific skills that an examinee has mastered, as well as overall competency level, may be obtained. The diagnostic IRT models relate scored attribute requirements in the items to performance. In chapter 3, Louis A. Roussos and his coauthors present an overview of diagnosing skills through diagnostic IRT models.

Another approach to model-based measurement of cognitive skills is based on multinomial processing tree (MPT) models. MPT models were initially developed to study normal cognition in specific experimental paradigms in cognitive psychology, but more recently MPT models have been applied to understand and measure how special populations differ in latent cognitive skills. In chapter 4, William Batchelder reviews the methodological issues that arise when MPT models are used for psychological assessment. He also describes recent applications to special populations.

In Part II, *Model-Based Approaches to Isolating Entangled Constructs*, two approaches are presented that have an interesting relationship to the traditional psychometric principle of internal consistency. Selecting items by internal consistency can lead to measuring either constructs at the wrong level of generality (Gustafsson, 2002) or trivial constructs (see Steinberg & Thissen, 1996). Model-based measurement approaches can aid in untangling the trivial or inappropriate constructs from the theoretically targeted construct. In chapter 5, Jan-Eric Gustafsson and Lisbeth Åberg-Bengtsson describe how model-based results reveal that the intended construct of a psychological test in itself can be understood in terms of both more general and more specific constructs. Such results have important implications for distinguishing between constructs because tests that fit unidimensional measurement models, in fact, typically reflect multiple