



Statistical Methods for Validation of Assessment Scale Data in Counseling and Related Fields

Dimiter M. Dimitrov



AMERICAN COUNSELING ASSOCIATION

5999 Stevenson Avenue

Alexandria, VA 22304

www.counseling.org



Statistical Methods for Validation of Assessment Scale Data in Counseling and Related Fields

Copyright © 2012 by the American Counseling Association. All rights reserved. Printed in the United States of America. 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, or stored in a database or retrieval system, without the written permission of the publisher.

10 9 8 7 6 5 4 3 2 1

American Counseling Association
5999 Stevenson Avenue
Alexandria, VA 22304
www.counseling.org

Director of Publications ♦ Carolyn C. Baker

Production Manager ♦ Bonny E. Gaston

Editorial Assistant ♦ Catherine A. Brumley

Copy Editor ♦ Kimberly W. Kinne

Text and cover design by Bonny E. Gaston.

Library of Congress Cataloging-in-Publication Data

Dimitrov, Dimitar M.

Statistical methods for validation of assessment scale data in counseling and related fields/

Dimitar M. Dimitrov.

p. cm.

Includes bibliographical references and index.

ISBN 978-1-55620-295-7 (alk. paper)

1. Counseling—Evaluation. 2. Social sciences—Statistical methods. I. Title.

BF636.6.D56 2012

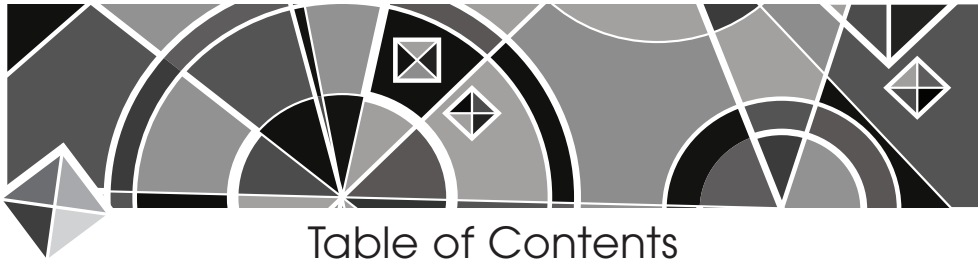
158'.30727—dc23

2011018595



To the memory of my parents,
Milko and Maria





Preface	vii
Acknowledgments	ix
About the Author	xi
 PART I ♦ Scales, Reliability, and Validity	
Chapter 1 Variables and Measurement Scales	3
1.1 Variables in Social and Behavioral Research	3
1.2 What Is Measurement?	4
1.3 Levels of Measurement	5
1.4 Typical Scales for Assessment in Counseling	7
1.5 Scaling	13
Summary	20
Chapter 2 Reliability	23
2.1 What Is Reliability?	23
2.2 Classical Concept of Reliability	24
2.3 Types of Reliability	28
2.4 Stratified Alpha	35
2.5 Maximal Reliability of Congeneric Measures	38
Summary	39
Chapter 3 Validity	41
3.1 What Is Validity?	41
3.2 Unified Construct-Based Model of Validity	42
Summary	50
Chapter 4 Steps in Instrument Development	53
4.1 Definition of Purpose	53
4.2 Instrument Specifications	54
4.3 Item Development	59
Summary	64
 PART II ♦ Factor Analysis in Construct Validation	
Chapter 5 Exploratory Factor Analysis	69
5.1 Correlated Variables and Underlying Factors	69
5.2 Basic EFA Models	70
5.3 The Principal Factor Method of Extracting Factors	73
5.4 Rotation of Factors	76
5.5 Some Basic Properties	79
5.6 Determining the Number of Factors	81

Table of Contents

5.7	Higher-Order Factors	86
5.8	Sample Size for EFA	87
5.9	Data Adequacy for EFA	87
5.10	EFA With Categorical Data	89
5.11	EFA in Collecting Evidence of Construct Validity	90
	Summary	91
Chapter 6	Confirmatory Factor Analysis	95
6.1	Similarities and Differences of EFA and CFA	95
6.2	CFA Model Specification	97
6.3	Dependent and Independent Variables in CFA	98
6.4	CFA Model Parameters	99
6.5	CFA Model Identification	100
6.6	Evaluation of CFA Model Adequacy	102
6.7	Factorial Invariance Across Groups	110
6.8	Testing for Factorial Invariance	112
6.9	Comparing Groups on Constructs	118
6.10	Higher-Order CFA	122
6.11	Points of Caution in Testing for Factorial Invariance	131
6.12	Sample Size for CFA	133
	Summary	134
Chapter 7	CFA-Based Models of Multitrait–Multimethod Data	143
7.1	Conventional MTMM Analysis	143
7.2	The Standard CFA Model	145
7.3	The CU Model	147
7.4	The CU–CFA Model	150
7.5	The Correlated Trait–Correlated Method Minus One [CTC($M - 1$)] Model	152
7.6	The Random Intercept Factor Model	156
7.7	The Hierarchical CFA (HCFA) Model	158
7.8	The Multilevel CFA (ML-CFA) Method	161
7.9	Conventional MTMM Analysis Using Latent Variable Modeling	165
7.10	Brief Guidelines for Selecting Models of MTMM Data	167
	Summary	169
	PART III ♦ Psychometric Scale Analysis	
Chapter 8	Conventional Scale Analysis	175
8.1	Analysis of Binary Scales	175
8.2	Analysis of Rating Scales	181
8.3	Estimation of Reliability for Congeneric Measures	186
	Summary	188
Chapter 9	Rasch-Based Scale Analysis	191
9.1	Rasch Model for Binary Data	191
9.2	Rating Scale Model (RSM)	200
	Summary	216
	References	219
	Index	249



Preface

The purpose of this book is to present statistical methods and procedures used in contemporary approaches to validation of targeted constructs through the use of assessment scales (tests, inventories, questionnaires, surveys, and so forth). An important clarification in this regard is that *validity* is a property of data and inferences made from data rather than a property of scales (or instruments in general). Although most references and examples are in the context of counseling, the methodology and practical know-how provided in this book directly apply to assessments in psychology, education, and other fields. The text is intended primarily for use by applied researchers, but it can also be useful to faculty and graduate students in their coursework, research, dissertations, and grants that involve development of assessment instruments and/or related validations.

To a large extent, the need for this book stemmed from my six-year work (2005–2011) as editor of *Measurement and Evaluation in Counseling and Development*, the official journal of the Association for Assessment in Counseling and Education, and as a reviewer for numerous professional journals in the areas of counseling, psychology, and education. In general, commonly occurring shortcomings in (mostly unpublished) manuscripts that deal with validation of assessment instruments relate to outdated conceptions of validity, lack of sound methodology, and/or problems with the selection and technical execution of statistical methods used to collect evidence about targeted aspects of validity. The approach to validation of assessment scale data and related statistical procedures presented in this book is based on the unified construct-based conception of validity (Messick, 1989, 1995), which is also reflected in the current *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). On the technical side, this book presents contemporary statistical methods and related procedures for evaluating psychometric properties of assessment scales. For example, exploratory and confirmatory factor analysis, testing for invariance of constructs across groups, multitrait–multimethod data analysis for validity evidence, and modern scale analysis are elaborated at both methodological and technical levels.

This book is organized in three parts comprising nine chapters. Part I (*Scales, Reliability, and Validity*) consists of four chapters. Chapter 1 presents variables and measurement scales, with focus on the nature of measurement, types of scales, and scaling procedures typical for assessment in the context of counseling, psychology, education, and other fields. Chapter 2 introduces the classical (true-score) model of score reliability, types of reliability, reliability of composite scores, and maximal reliability. Chapter 3 presents the unified construct-based model of validity (Messick, 1989, 1995). Chapter 4 outlines major steps in the development of an assessment instrument within the framework of the adopted validity model.

Part II (*Factor Analysis in Construct Validation*) consists of three chapters. Chapter 5 deals with exploratory factor analysis—a brief introduction of the EFA framework, contemporary approaches to determining the number of factors, and issues of sample size, data

adequacy, and categorical data. Chapter 6 deals with confirmatory factor analysis (CFA). As this chapter plays a central role under the conception of validity adopted in the book, topics of critical importance such as CFA model–data fit, evaluation of model adequacy, and testing for factorial invariance of (first- and higher-order) CFA models are addressed with methodological and technical details in the context of construct validation. Chapter 7 presents a variety of CFA-based models of multitrait–multimethod data analysis for collecting convergent and discriminant evidence, as well as evidence of method bias, as related to the external aspect of construct validity.

Part III (*Psychometric Scale Analysis*) consists of two chapters. Chapter 8 deals with classical scale analysis of binary and rating scales, with a focus on procedures that can be useful to researchers in piloting stages of development and validation of an assessment instrument. Chapter 9 presents Rasch-based analysis of binary and rating scales, and particular attention is paid to optimizing the effectiveness of rating scales by addressing issues of disordering in rating scale categories and their thresholds, person–item distribution mapping, and dimensionality of assessment measures.

From a pedagogical perspective, the presentation of topics was guided by the intent to provide applied researchers with understandable treatment of contemporary statistical methods and procedures that they would be able to apply in development and validation of assessment scale data. The hope is that this goal is achieved by minimized use of mathematical symbols and formulas and focus on conceptual understanding of methods and procedures, underlying assumptions, possible pitfalls, and common misconceptions. This strategy is enhanced by the use of numerous illustrative examples, tables, and figures throughout the text. Practical applications of relatively complex procedures are facilitated by the inclusion of operationalized (step-wise) guidance for their implementation and computer code in Mplus (Muthén & Muthén, 2008). Of course, given the description of such procedures, they can be translated into computer source codes for other popular software packages such as LISREL, EQS, or Amos.



Acknowledgments

I would like to thank all colleagues, friends, and family members for their encouragement and support during my work on this book. I truly appreciate the guidance, expertise, and support provided by Carolyn Baker, director of publications for the American Counseling Association (ACA), from the initial idea about the need for such a book to its final publication. I am also grateful to the supportive role of the ACA Publications Committee.

I would like to acknowledge the expertise and contribution of the reviewers, Everett V. Smith, Jr.; Thomas J. Smith; Carolyn Baker; and Catherine Y. Chang, all of whom provided valuable comments and suggestions on improving the book. I am also grateful to my family for their patience and encouragement during the time this book was written.

—*Dimitar M. Dimitrov*
George Mason University



Dimiter M. Dimitrov, PhD, is professor of educational measurement and statistics in the Graduate School of Education at George Mason University in Fairfax, Virginia. He earned his bachelor's degree in mathematics and a PhD in mathematics education from the University of Sofia, Bulgaria, in 1984 as well as a PhD in educational psychology from Southern Illinois University at Carbondale in 1995. His teaching experience includes courses on multivariate statistics, quantitative research methods, modern measurement, generalizability theory, and structural equation modeling. Dr. Dimitrov's professional work—which has resulted in numerous journal articles, books, and book chapters—has received national and international recognition. He has served as president of the Mid-Western Educational Research Association (2008–2009), program chair of the SIG Rasch Measurement of the American Educational Research Association, and editor of *Measurement and Evaluation in Counseling and Development*, the official journal of the Association for Assessment in Counseling and Education (2005–2011). Dr. Dimitrov has also lectured on modern measurement and latent variable modeling at universities in Russia and Spain. He has served on the editorial board of prestigious professional journals such as *Educational Researcher*, *Educational and Psychological Measurement*, *Journal of Applied Measurement*, and *Research in the Schools*. Dr. Dimitrov is multilingual and has lectured and published professional work in English, Bulgarian, Russian, and French.

His email address is: ddimitro@gmu.edu.

