

ASSESSING MEASUREMENT INVARIANCE FOR APPLIED RESEARCH

Assessing Measurement Invariance for Applied Research will provide psychometricians and researchers across diverse disciplines in the social sciences with the necessary knowledge and skills to select and apply appropriate methods to assess measurement invariance. It is a user-friendly guide that describes a variety of statistical methods using a pedagogical framework emphasizing conceptual understanding, with extensive illustrations that demonstrate how to use software to analyze real data. A companion website (people.umass.edu/cswells) provides downloadable computer syntax and the data sets demonstrated in this book so readers can use them to become familiar with the analyses and understand how to apply the methods with proficiency to their own work. Evidence-supported methods that can be readily applied to real-world data are described and illustrated, providing researchers with many options from which to select given the characteristics of their data. The approaches include observed-score methods and those that use item response theory models and confirmatory factor analysis.

DR. CRAIG S. WELLS is a professor in the Research, Educational Measurement and Psychometrics program at the University of Massachusetts (UMass) Amherst, where he teaches courses on statistics and psychometrics, including structural equation modeling and psychometric modeling. He is also the associate director for the Center for Educational Assessment at UMass. His research interests pertain to statistical methods for detecting differential item functioning, and assessing the fit of item response theory models. He also has a keen interest in philosophy of science, especially how it relates to hypothesis testing. He coedited *Educational Measurement: From Foundations to Future* (2016) and was president of the Northeastern Educational Research Association in 2017.

EDUCATIONAL AND PSYCHOLOGICAL TESTING
IN A GLOBAL CONTEXT

Editor

Neal Schmitt, *Michigan State University*

The Educational and Psychological Testing in a Global Context series features advanced theory, research, and practice in the areas of international testing and assessment in psychology, education, counseling, organizational behavior, human resource management, and all related disciplines. It aims to explore, in great depth, the national and cultural idiosyncrasies of test use and how they affect the psychometric quality of assessments and the decisions made on the basis of measures. Our hope is to contribute to the quality of measurement and to facilitate the work of professionals who must use practices or measures with which they may be unfamiliar or adapt familiar measures to a local context.

Published Titles:

DRAGOȘ ILIESCU,
Adapting Tests in Linguistic and Cultural Contexts, 2017

JOHN C. SCOTT, DAVE BARTRAM and DOUGLAS H. REYNOLDS (eds.),
International Applications of Web-Based Testing: Challenges and Opportunities,
2017

WILLIAM H. SCHMIDT et al.,
Schooling Across the Globe: What We Have Learned from 60 Years of
Mathematics and Science International Assessments, 2018

MARIA ELENA OLIVERI and CATHY WENDLER,
Higher Education Admission Practices: An International Perspective, 2020

Forthcoming Titles:

SUMAYA LAHER,
Comparative Histories of Psychological Assessment

ASSESSING MEASUREMENT INVARIANCE FOR APPLIED RESEARCH

CRAIG S. WELLS

University of Massachusetts Amherst



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press
978-1-108-48522-7 — Assessing Measurement Invariance for Applied Research
Craig S. Wells
Frontmatter
[More Information](#)

CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom
One Liberty Plaza, 20th Floor, New York, NY 10006, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India
79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.
It furthers the University's mission by disseminating knowledge in the pursuit of
education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108485227
DOI: 10.1017/9781108750561

© Cambridge University Press 2021

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2021

A catalogue record for this publication is available from the British Library.

ISBN 978-1-108-48522-7 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of
URLs for external or third-party internet websites referred to in this publication
and does not guarantee that any content on such websites is, or will remain,
accurate or appropriate.

Cambridge University Press
978-1-108-48522-7 — Assessing Measurement Invariance for Applied Research
Craig S. Wells
Frontmatter
[More Information](#)

*This book is dedicated to my loving family:
Amanda, Madeline, and Elianna.*

Cambridge University Press

978-1-108-48522-7 — Assessing Measurement Invariance for Applied Research

Craig S. Wells

Frontmatter

[More Information](#)

Contents

<i>List of Figures</i>	<i>page</i> ix
<i>List of Tables</i>	xii
<i>Preface</i>	xix
1 Introduction	1
What Is Measurement Invariance?	1
Why Should We Assess Measurement Invariance?	5
Forms of DIF	8
Classification of DIF Detection Methods	10
The Conditioning Variable: To Purify or Not to Purify?	11
Considerations When Applying Statistical Tests for DIF	13
2 Observed-Score Methods	17
Transformed Item Difficulty (Delta-Plot) Method	19
Mantel–Haenszel DIF Procedure	31
Standardization DIF Procedure	52
Logistic Regression	65
SIBTEST	86
Applying Observed-Score Methods to Incomplete Data Sets	104
3 Item Response Theory	108
Unidimensional IRT Models	111
Dichotomous Item Responses	112
Polytomous Item Responses	120
Multidimensional IRT Models	127
Model Assumptions	129
Parameter Estimation	136
The Latent Variable Scale	141
Linking IRT Scales	143
Item and Testing Information Functions	146
Test Information	149
IRT Parameter Estimation Software	152

viii	<i>Contents</i>	
4	Methods Based on Item Response Theory	161
	<i>b</i> - and <i>a</i> -Plots	166
	Lord's Chi-Square and Wald Statistic	177
	Likelihood-Ratio Test	198
	Raju's Area DIF Measures	213
	DFIT	222
	Applying IRT DIF Methods to Incomplete Data Sets	243
5	Confirmatory Factor Analysis	245
	Basic Principles of CFA	246
	Example of CFA Using Mplus	270
	Analyzing Categorical Indicators	281
	Description and Example of the Bifactor Model	288
6	Methods Based on Confirmatory Factor Analysis	295
	Multigroup Confirmatory Factor Analysis	296
	MIMIC Model	332
	Applying CFA-Based Methods to Incomplete Data Sets	367
	<i>Appendix A: A Brief R Tutorial</i>	369
	<i>References</i>	376
	<i>Author Index</i>	388
	<i>Subject Index</i>	391

You can download the data files and computer syntax
described in this book at the companion website:
people.umass.edu/cswells

Figures

1.1 An example of functional relationships for two groups on a dichotomously scored item that is invariant.	<i>page 3</i>
1.2 An example of functional relationships for two groups on a dichotomously scored item that lacks invariance.	4
1.3 An illustration of an invariant and non-invariant polytomous item.	6
1.4 Two examples of nonuniform DIF.	9
2.1 Plot of the Δ -values for the reference and focal groups.	20
2.2 Delta-plot of the reference and focal groups produced by the <code>diagPlot</code> function within R package <code>deltaPlotR</code> .	26
2.3 Plot of the conditional p -values for the reference and focal groups: (a) item 10 and (b) item 20.	54
2.4 Plot of the conditional p -values and difference in conditional p -values across the matching variable for item 20.	61
2.5 Plot of conditional expected scores and difference in conditional expected scores across the matching variable for item 11.	63
3.1 An example of an IRF for a hypothetical multiple-choice math item.	109
3.2 An example of an IRF from the 3PL model.	113
3.3 Examples of IRFs from the 3PL model.	115
3.4 Examples of IRFs from the 2PL model.	118
3.5 Examples of IRFs from the 1PL model.	119
3.6 Example of boundary response functions for Samejima's GR model.	122
3.7 Category response functions based on the boundary response functions shown in Figure 3.6.	123
3.8 Conditional expected item score based on the category response functions shown in Figure 3.7.	124

3.9	Category response functions for two partial-credit items from the GPC model.	126
3.10	An example of an IRS for a multidimensional 2PL model.	129
3.11	An example of an empirical and model-based IRF for a fitting item (top) and a misfitting item (bottom).	135
3.12	An example of the conditional observed proportions for an item with 10 intervals (groups) used in item parameter estimation.	137
3.13	An example of a TCF based on the 3PL model for a 36-item, multiple-choice test.	142
3.14	A scatterplot of unscaled item difficulty parameter estimates for two groups with disparate proficiency distributions.	144
3.15	Examples of the IIF for a 2PL and 3PL model item.	148
3.16	Examples of the IIF from the 3PL model.	150
3.17	Example of a TIF and CSEE based on the 3PL model for a 36-item test.	151
3.18	A plot for the conditional reliability along with the TIF from Figure 3.17.	152
4.1	IRFs illustrating uniform and nonuniform DIF.	163
4.2	IRFs illustrating uniform DIF where the difference in the difficulty parameter estimates equals 0.5 for a high and low discriminating item.	165
4.3	<i>b</i> -plot using difficulty parameter estimates from the 3PL model.	167
4.4	<i>a</i> -plot using discrimination parameter estimates from the 3PL model.	172
4.5	<i>b</i> -plot from Samejima's GR model using the thresholds.	174
4.6	<i>b</i> -plot from Samejima's GR model using the average of the thresholds.	175
4.7	<i>b</i> -plot using <i>bs</i> from the 3PL model and the individual thresholds from the GR model.	177
4.8	Example of noncentral chi-square distribution used to test a range-null hypothesis using Lord's chi-square statistic.	182
4.9	Comparison of IRFs for the reference and focal groups for item 20, flagged as DIF using Lord's chi-square test.	189
4.10	Comparison of expected item scores for item 11, which was flagged for DIF using Lord's chi-square test.	194
4.11	Plot of IRFs for the focal and reference groups produced by the <code>PlotNcdif</code> function for item 20 (large DIF).	234

List of Figures

xi

4.12 Plot of IRFs for the focal and reference groups produced by the <code>PlotNcdif</code> function for item 9 (trivial DIF).	235
4.13 Plot of expected test scores for the focal and reference groups produced by the <code>plot.tcf</code> function.	237
4.14 Plot of IRFs for the focal and reference groups produced by the <code>PlotNcdif</code> function for item 4.	241
4.15 Plot of IRFs for the focal and reference groups produced by the <code>PlotNcdif</code> function for item 11.	242
5.1 Path diagram for a two-factor model.	248
5.2 Illustrating under-identified, just-identified, and over-identified models using simple linear regression.	252
5.3 Illustration of thresholds for an ordinal item with four categories.	283
5.4 Example of a bifactor model.	289
6.1 Example of a MIMIC model to test uniform DIF in the first indicator.	332
6.2 Example of a MIMIC model that includes the interaction term.	334
6.3 Path diagram illustrating longitudinal measurement invariance.	360
A.1 Screenshot of the R website.	370
A.2 Screenshot of the R window and console.	371

Tables

2.1	Selected Output from Delta-Plot Analysis Using the <code>deltaPlot</code> Function in R and a Threshold of 1.5 to Flag Items as DIF.	<i>page</i> 25
2.2	2×2 Contingency Table for the m th Score Level.	32
2.3	Hypothetical 2×2 Contingency Table Data for the m th Score Level.	32
2.4	Abbreviated Output from MH DIF Analysis Using the <code>difMH</code> Function in R.	40
2.5	Output from the <code>CI.for.D</code> Function that Provides the $100(1-\alpha)\%$ Confidence Interval for $D_{\alpha_{MH}}$.	42
2.6	Output from the <code>PDIF</code> Function that Computes P_{DIF} and Its Corresponding $100(1-\alpha)\%$ Confidence Interval.	44
2.7	$G \times 2$ Contingency Table for the m th Score Level.	45
2.8	Output (Abbreviated) from GMH DIF Analysis Using the <code>difGMH</code> Function in R.	47
2.9	$2 \times C$ Contingency Table for the m th Score Level.	49
2.10	Output from Mantel DIF Procedure for Polytomous Data using the <code>Mantel.poly</code> Function in R.	52
2.11	Output from GMH DIF Procedure for Polytomous Data Using the <code>GMH.poly</code> Function in R.	53
2.12	Abbreviated Output from Standardization DIF Analysis Using the <code>difSTD</code> Function in R.	59
2.13	Output from Standardization DIF Analysis for Polytomous Data Using the <code>SMD</code> Function in R.	64
2.14	Output from the <code>lrm</code> Function When Performing Logistic Regression for Model 3 that Includes the Matching and Grouping Variables along with the Interaction Term for Item 1.	72

List of Tables

xiii

2.15	Output from the <code>lrm</code> Function When Performing Logistic Regression for Model 1 that Includes Only the Matching Variable for Item 1.	73
2.16	Abbreviated Output from the <code>DIF.Logistic</code> Function.	75
2.17	Output from the <code>lrm</code> Function When Performing Logistic Regression for Model 3 for Dichotomous Data and Three Groups for Item 4.	78
2.18	Output from the <code>lrm</code> Function When Performing Logistic Regression for Model 1 for Dichotomous Data and Three Groups for Item 4.	79
2.19	Abbreviated Output from the <code>DIF.Logistic.MG</code> Function.	81
2.20	Output from the <code>lrm</code> Function When Performing Logistic Regression for Model 3 for Polytomous Data for Item 1.	84
2.21	Output from the <code>lrm</code> Function When Performing Logistic Regression for Model 1 for Polytomous Data for Item 1.	85
2.22	Abbreviated Output from the <code>DIF.Logistic</code> Function.	87
2.23	Output from the <code>SIBTEST</code> Function for Testing Item 21 for DIF.	95
2.24	Conditional Item Scores from the <code>SIBTEST</code> Function for Testing Item 21 for DIF.	96
2.25	Abbreviated Output from <code>SIBTEST.exp</code> Function Applied to Dichotomous Data.	98
2.26	Output from the <code>SIBTEST</code> Function for Testing Item 12 for DIF.	100
2.27	Output from <code>SIBTEST.exp</code> Function Applied to Polytomous (Likert-type) Data with the Two-Step Purification Procedure.	102
2.28	Output from the <code>SIBTEST</code> Function for Testing DBF Based on Items 16 through 20.	103
3.1	<code>flexMIRT</code> Input File for Fitting 3PL and GPC Models.	154
3.2	Item Parameter Estimates for 3PL and GPC Model from <code>flexMIRT</code> .	157
3.3	<code>flexMIRT</code> Input File for Fitting Samejima's GR Model.	158
3.4	Item Parameter Estimates for GR Model from <code>flexMIRT</code> Reported in Slope-Intercept Form (Top) and IRT Parameterization.	159
4.1	Output from the <i>b</i> -plot Analysis Using the <code>b.plot</code> Function in R with the Two-Step Purification Procedure.	170
4.2	Metadata for the Reference Group Found in <code>Ref.est</code> .	184

xiv	<i>List of Tables</i>	
4.3	Abbreviated Output from the <code>LordChi</code> Function for Testing DIF under the 3PL Model.	186
4.4	Metadata for the Reference Group's Likert-type Data in the Data Frame <code>Ref.est</code> .	191
4.5	Abbreviated Printout of <code>Lord.results</code> for Likert Data Using Samejima's GR Model.	193
4.6	Metadata for the Reference Group's Stored in the Data Frame <code>Ref.est</code> .	196
4.7	Example Data File (Abbreviated) with 30 Dichotomous Items (Columns 1 through 30) and Group Membership in Column 31.	202
4.8	Information to Execute the IRTLRDIF Program for the 3PL Model.	204
4.9	Abbreviated Output from the IRTLRDIF Program.	205
4.10	Content of Input File to Run with IRTLRDIF.	206
4.11	Content of Input File to Run with IRTLRDIF with an Anchor.	208
4.12	Content of Input File to Run with IRTLRDIF for Samejima's GR Model.	208
4.13	Output from IRTLRDIF for Testing DIF Using Samejima's GR Model.	209
4.14	Content of Input File to Run with IRTLRDIF for a Mixed-Format Test that Uses the 3PL, 2PL, and Samejima's GR Models.	211
4.15	Abbreviated Output from the <code>difRaju</code> Function for the 3PL Model with g Fixed to 0.2.	218
4.16	Abbreviated Output from <code>difRaju</code> Function for the 1PL Model.	221
4.17	<i>NCDIF</i> Estimates from the 3PL Model.	230
4.18	<i>CDIF</i> Estimates for the 3PL Model.	236
4.19	<i>NCDIF</i> Estimates from Samejima's GR Model.	239
5.1	Hypothetical Correlation Matrix for Six Observed Variables Based on Two Factors.	247
5.2	Hypothetical Correlation Matrix for Six Observed Variables Based on One Factor.	250
5.3	Mplus Input Code for Two-Factor CFA Model.	271
5.4	Descriptive Statistics for Subscore Indicators Reported by Mplus.	273
5.5	Model Fit Information for Fitting a Two-Factor CFA Model Reported by Mplus.	275

List of Tables

xv

5.6	Standardized Residuals from a Two-Factor Model Reported by Mplus.	276
5.7	Unstandardized Parameter Estimates.	277
5.8	Standardized Parameter Estimates.	278
5.9	R-Squared Values.	279
5.10	Mplus Input Code for Fitting a One-Factor Model to the Subscores.	279
5.11	Mplus Input Code for Fitting a Two-Factor Model to Item Responses.	284
5.12	Mplus Input Code for Fitting a One-Factor Model to Item Responses.	286
5.13	Mplus Output for Chi-Square Difference Test.	287
5.14	Mplus Input Code for Fitting a One-Factor Model to Likert-Type Responses.	287
5.15	Mplus Input Code for Fitting a Bifactor Model.	290
5.16	Parameter Estimates from the Bifactor and One-Factor Models.	291
6.1	Mplus Input File for Testing Configural Invariance (Equal Dimensionality Structure) for the Reference Group (Reference = 0) on the Subscores.	304
6.2	Fit Statistics for Evaluating Configural, Metric, Scalar, and Strict Factorial Invariance on the Subscore Data.	306
6.3	Mplus Input File for Configural Invariance Model in Which the Factor Loadings and Intercepts Are Freely Estimated in Each Group.	307
6.4	Parameter Estimates from Mplus for the Configural Invariance, Metric Invariance, Scalar Invariance, and Partially Invariant Scalar Models.	308
6.5	Mplus Input File for Metric Invariance Model in Which the Factor Loadings Are Constrained to Be Equal between the Groups While the Intercepts Are Freely Estimated.	309
6.6	Mplus Input File for the Scalar Invariance Model in Which the Factor Loadings and Intercepts Are Constrained to be Equal between the Groups.	310
6.7	Modification Indices Reported in the Mplus Output for the Scalar Model.	311
6.8	Mplus Input File for the Scalar Partial Invariance Model in Which the Intercept for the Fifth Indicator (C5) Has Been Freely Estimated in Both Groups.	312

6.9	Modification Indices Reported in the Mplus Output for the Partially Invariant Scalar Model.	313
6.10	Mplus Input File for Testing Strict Factorial Invariance in Which the Factor Loadings, Intercepts (Except for C5), and Indicator Residuals Have Been Constrained to be Equal between the Groups.	314
6.11	Mplus Input File for Configural, Metric, and Scalar Invariance Using the MODEL = CONFIG METRIC SCALAR Command.	315
6.12	Abbreviated Mplus Output Reporting the Statistics for Testing Configural, Metric, and Scalar Invariance.	316
6.13	Mplus Input File for Testing Configural Invariance (Equal Dimensionality Structure) for the Reference Group (Reference = 0) on the Item Responses Using Categorical Indicators.	318
6.14	Fit Statistics for Evaluating Configural, Metric, Scalar, and Strict Factorial Invariance on the Subscore Data.	320
6.15	Mplus Input File for Configural Invariant (Baseline) Model in Which Factor Loadings and Thresholds Are Freely Estimated in Each Group Using Categorical Indicators.	321
6.16	Mplus Input File for the Metric Invariance Model in Which Factor Loadings Are Constrained to Be Equal between Groups While the Intercepts Are Freely Estimated Using Categorical Indicators.	323
6.17	Mplus Input File for the Partial Metric Invariance Model in Which the Factor Loadings Are Constrained to Be Equal between Groups Except for Indicator 19.	324
6.18	Mplus Input File for the Scalar Invariance Model in Which the Factor Loadings and Intercepts Are Constrained to Be Equal between the Groups Using Categorical Indicators.	326
6.19	Mplus Input File for the Scalar Partial Invariance Model in Which the Thresholds for the 19th and 40th Indicators Have Been Freely Estimated in Both Groups.	328
6.20	Threshold Estimates for Items 12 and 40.	328
6.21	Mplus Input File for the Scalar Partial Invariance Model in Which the Thresholds for the 12th, 19th, and 40th Indicators Have Been Freely Estimated in Both Groups.	329
6.22	Mplus Input File for the Strict (Partial) Invariance Model in Which the Residual Variances Are Unconstrained.	330

List of Tables

xvii

6.23 Mplus Input File for Testing a One-Factor Model on the Subscore Data.	338
6.24 Mplus Input File for Testing Whether the First Content Domain Functions Differentially between the Groups.	339
6.25 Mplus Input File for Fixing the Path Coefficients between the Grouping Variable and Indicators to Zero.	340
6.26 Mplus Input File for Testing Nonuniform and Uniform DIF.	343
6.27 Mplus Input Code for Fixing the Path Coefficients to Zero and Requesting Modification Indices when Testing Nonuniform and Uniform DIF.	344
6.28 Mplus Input File for Testing Uniform DIF via the Wald Statistic and Estimating the Parameters Using Robust WLS Estimation.	348
6.29 Augmented Model Where the Path Coefficient for the Grouping Variable Was Estimated for the First Indicator.	349
6.30 Compact Model Where the Path Coefficient for the Grouping Variable Was Fixed to Zero for Each Indicator.	349
6.31 Mplus Input Code When Fixing the Path Coefficients to Zero.	351
6.32 Logistic Regression Odds Ratio Statistics Reported in Mplus Output.	352
6.33 Mplus Input File for Testing Nonuniform and Uniform DIF Using the Categorical Indicators.	353
6.34 Mplus Input Code for Fixing the Path Coefficients to Zero for Categorical Indicators.	355
6.35 Mplus Input Code for Estimating the Path Coefficient Regressing Item 12 on the Grouping Variable but Fixing the Path Coefficient to Zero for the Interaction Term.	357
6.36 Mplus Input Code for Fitting a Baseline Model for Assessing Longitudinal Measurement Invariance.	362
6.37 Fit Statistics for Evaluating Configural, Metric, Scalar, and Strict Factorial Longitudinal Invariance for the Anxiety Test.	364
6.38 Mplus Input Code for Fitting the Metric Longitudinal Invariance Model.	365
6.39 Input Code for Fitting the Scalar Longitudinal Invariance Model.	366
A.1 R Code Used in This Appendix.	372

Cambridge University Press

978-1-108-48522-7 — Assessing Measurement Invariance for Applied Research

Craig S. Wells

Frontmatter

[More Information](#)

Preface

Why write a book about assessing measurement invariance? This question is not about justifying the importance of examining measurement invariance (I address that issue in Chapter 1) but instead about why a book on measurement invariance is useful for social science researchers. In the past 30 years, there have been several excellent texts that have described statistical methods for detecting a lack of measurement invariance (often referred to as differential item functioning), such as Holland and Wainer (1993), Camilli and Shepard (1994), Zumbo (1999), Osterlind and Everson (2009), Engelhard (2012), and Millsap's comprehensive *Statistical Approaches to Measurement Invariance* (2011) in which he provided a unifying theory of measurement invariance. Each of these books provided a valuable contribution to the literature, and we are indebted to the authors and the many other researchers who have added to the theory and practice of assessing measurement invariance. What sets this book apart from those excellent resources on this topic is that, as the title implies, this book focuses on the practical application of assessing measurement invariance, with less emphasis on theoretical development or exposition. As such, a primary emphasis of the book is to describe the methods using a pedagogical framework, followed by extensive illustrations that demonstrate how to use software to analyze real data. My intention is that you will use this book to learn about practical methods to assess measurement invariance and how to apply them to your own data. This book is intended to be a user-friendly guide for anyone who wants to assess measurement invariance in their own work. To support this goal, the computer syntax and data sets used in this book are available for download at the following website: www.people.umass.edu/cswells.

As someone who has taught statistics and psychometrics for many years, I have learned that students learn better when they are actively engaged in the learning process. Therefore, I invite you to download the example data files and syntax and perform the analyses while you read along in this

book. My goal is to provide you with enough instruction and practice in the most useful methods so that you can apply what you have learned to novel problems and data sets in your own work.

The audience I envisioned while writing this book is applied researchers, both graduate students and professionals, from the social sciences in which educational tests, psychological inventories, questionnaires, and surveys are administered. This includes researchers from diverse fields, such as education, psychology, public policy, higher education, marketing, sports management, sociology, and public health, to name a few. The purpose of this book is to provide researchers and psychometricians from these diverse fields with the necessary knowledge and skills to select and apply an appropriate method to assess measurement invariance for the problems they are trying to solve and the research questions they are trying to answer. Regardless of your field, I wrote this book assuming you understand basic statistics taught in an introductory statistics course (e.g., mean, standard deviation, correlation, regression, effect size, hypothesis testing, confidence intervals), as well as the concepts of reliability and validity.

This book includes chapters that describe the big ideas and the theoretical concepts of measurement invariance and various statistical procedures, while other chapters include a pedagogical framework for you to learn a new method, including guided practice opportunities. In Chapter 1 I describe the basic ideas of measurement invariance, including why it is important to assess, and I address important issues to consider when applying the methods described in this book. Chapters 2, 4, and 6 are written to teach you specific methods for assessing measurement invariance. The basic structure of each of these chapters is a description of the method, followed by detailed illustrations on how to examine measurement invariance using computer software. Chapter 2 describes observed-score methods where the groups are matched using raw scores. Chapter 4 describes detection methods that use item response theory (IRT) models. Chapter 6 describes methods that rely on confirmatory factor analysis (CFA). For those who are unfamiliar with IRT, Chapter 3 provides a description of the basic concepts of and models used in IRT. I also provide an illustration of how to use the computer program flexMIRT (Cai, 2017) to estimate the parameters for several IRT models. Chapter 5 provides a description of the underlying concepts of CFA that are needed to understand the CFA-based methods described in Chapter 6. Several CFA models are fit to data using the computer program Mplus (Muthén & Muthén, 2008–2017). Finally, Appendix A provides a brief tutorial on the

Preface

xxi

computer program R, because many of the techniques described in this book rely on packages in R.

There are many statistical methods available that I could have included in this book. Given that I was writing for practitioners and not writing a methodological tome that addresses all methods that have been created, I had to strategically select the methods I wanted to address. To accomplish this goal, I decided to include methods that met four criteria. The methods must (1) have been studied empirically and shown to provide reasonably accurate assessment of measurement invariance; (2) be able to incorporate an effect size or some type of information for determining whether the lack of invariance was nontrivial; (3) be relatively simple to use, especially for practitioners who are non-psychometricians; and (4) have software readily available to implement them. Indeed, there are methods that meet these criteria that I may not have included, but there will always be more to learn and write.

One of the challenges in writing a book that encompasses methods from different areas (i.e., observed-score and IRT- and CFA-based detection methods) is that the same symbols can take on different meanings depending on the context. I have attempted to use the symbols in their original meaning, which may cause confusion. For example, the Greek letter α is often used to represent the nominal Type I error rate, common odds ratio, or item discrimination parameter in an IRT model. It is important for the reader to discern the appropriate meaning of a symbol based on the context. To help prevent possible confusion, I have tried to define the symbols when they are first introduced in a chapter.

This book would not have been written if it were not for the support and encouragement of many people. First, I would like to acknowledge my colleagues at UMass Amherst: Steve Sireci, who recommended I write the book and whose support, guidance, and mentoring has had a positive impact on my career; Ronald Hambleton, whose ideas have permeated my understanding of measurement invariance and psychometrics in general; April Zenisky, Lisa Keller, Jennifer Randall, and Scott Monroe, who have all supported me when I needed it. I also would like to acknowledge Allan Cohen, Daniel Bolt, James Wollack, and Michael Subkoviak for their tutelage and persistent support during my doctoral studies; and Ron Serlin, who taught me the value of philosophy of science and the range-null hypothesis, which I have incorporated into DIF detection methods. I would also like to thank Neal Schmitt for giving me the opportunity to write this book, as well as for providing encouraging words along the journey. I would be remiss if I did not acknowledge the hard work and

support of Emily Watton, my editor at Cambridge University Press, for her tireless effort in bringing this manuscript to press, and Clare Diston, my copyeditor, whose attention to detail and knowledge for the English language brought out the best in my writing. Last but not least, I would like to express special thanks to my wife, Amanda, for her encouragement, support, and valuable feedback.