

CONTENTS

Preface	xiii
The Conceptual Orientation of This Book, Its Purpose, and the Intended Audience	xiii
Organizational Overview	xiv
New to This Edition	xvi
General Changes	xvi
Chapter-Specific Changes	xviii
Author's Acknowledgments	xxi
Publisher's Acknowledgments	xxii
About the Author	xxiv
Chapter 1 • Psychometrics and the Importance of Psychological Measurement	1
Why Psychological Testing Matters to You	2
Observable Behavior and Unobservable Psychological Attributes	4
Psychological Tests: Definition and Types	7
What Is a Psychological Test?	7
Types of Tests	8
What Is Psychometrics?	11
Psychometrics	11
A Brief History of Psychometrics	11
Challenges to Measurement in Psychology	13
The Importance of Individual Differences	18
But Psychometrics Goes Well Beyond "Differential" Psychology	19
Suggested Readings	20
PART I • BASIC CONCEPTS IN MEASUREMENT	21
Chapter 2 • Scaling	23
Fundamental Issues With Numbers	24
The Property of Identity	25
The Property of Order	26
The Property of Quantity	27
The Number 0	28
Units of Measurement	30
Additivity and Counting	32
Additivity	32
Counts: When Do They Qualify as Measurement?	34

Four Scales of Measurement	35
Nominal Scales	35
Ordinal Scales	36
Interval Scales	37
Ratio Scales	38
Scales of Measurement: Practical Implications	39
Additional Issues Regarding Scales of Measurement	40
Technical Appendix: R Syntax	41
Summary	46
Suggested Readings	47
Chapter 3 • Differences, Consistency, and the Meaning of Test Scores	49
The Nature of Variability	49
Importance of Individual Differences	50
Variability and Distributions of Scores	52
Central Tendency	54
Variability	54
Distribution Shapes and Normal Distributions	58
Quantifying the Association or Consistency Between Distributions	61
Interpreting the Association Between Two Variables	61
Scatterplots: Visually Representing the Association	
Between Two Variables	62
Covariance	64
Correlation	68
Variance and Covariance for "Composite Variables"	69
Binary Items	71
Interpreting Test Scores	74
Needed: An Interpretive Frame of Reference	75
z Scores (Standard Scores)	77
Converted Standard Scores (Standardized Scores)	81
Percentile Ranks	82
Normalized Scores	85
Test Norms	86
Representativeness of the Reference Sample	87
Technical Appendix: R Syntax	88
Summary	94
Suggested Readings	95
Chapter 4 • Test Dimensionality and Factor Analysis	97
Test Dimensionality	99
Three Dimensionality Questions: What They Are and Why They Matter	100
Unidimensional Tests	101
Multidimensional Tests With Correlated Dimensions (Tests With Higher-Order Factors)	103

Multidimensional Tests With Uncorrelated Dimensions	105
The Psychological Meaning of Test Dimensions	106
Factor Analysis: Examining the Dimensionality of a Test	107
The Logic and Purpose of Exploratory Factor Analysis:	
A Conceptual Overview	107
Conducting and Interpreting an Exploratory Factor Analysis	110
A Deeper Perspective on Factors, Factor Loadings, and Rotation	126
Factor Analysis of Binary Items	132
A Quick Look at Confirmatory Factor Analysis	133
Technical Appendix: R Syntax	134
Summary	139
Suggested Readings	140
PART II • RELIABILITY	141
Chapter 5 • Reliability: Conceptual Basis	143
Overview of Reliability and Classical Test Theory	145
Observed Scores, True Scores, and Measurement Error	147
Variances in Observed Scores, True Scores, and Error Scores	150
Four Ways to Think of Reliability	154
Reliability as the Ratio of True Score Variance to	
Observed Score Variance	155
Reliability as Lack of Error Variance	157
Reliability as the (Squared) Correlation Between	
Observed Scores and True Scores	159
Reliability as the Lack of (Squared) Correlation	
Between Observed Scores and Error Scores	161
Reliability and the Standard Error of Measurement	163
From Theory to Practice: Measurement Models and	
Their Implications for Estimating Reliability	165
Overview of Key Assumptions	166
Parallel Tests	170
Tau-Equivalent and Essentially Tau-Equivalent Tests	173
Congeneric Tests	176
Tests With Correlated Errors	177
Summary	178
Domain Sampling Theory	178
Summary	179
Suggested Readings	180
Chapter 6 • Empirical Estimates of Reliability	181
Alternate Forms Method of Estimating Reliability	183
Test-Retest Method of Estimating Reliability	186
Internal Consistency Method of Estimating Reliability	190
Split-Half Estimates of Reliability	191
"Raw" Coefficient Alpha	195

"Standardized" Coefficient Alpha	201
Raw Alpha for Binary Items: KR_{20}	203
Omega	205
On the Assumptions Underlying Alpha and Omega, the Relative Applicability of Those Indices, and Their Limitations	205
Internal Consistency Versus Dimensionality	209
Factors Affecting the Reliability of Test Scores	209
Sample Heterogeneity and Reliability Generalization	216
Reliability of Difference Scores	217
Defining Difference Scores	218
Estimating the Reliability of Difference Scores	220
Factors Affecting the Reliability of Difference Scores	221
The Problem of Unequal Variability	222
Difference Scores: Summary and Caution	226
Technical Appendix: R Syntax	228
Summary	233
Suggested Readings	234
Note	235
Chapter 7 • The Importance of Reliability	237
Applied Behavioral Practice: Evaluation of an Individual's Test Score	237
Point Estimates of True Scores	238
Confidence Intervals	242
Debate and Alternatives	244
Summary	245
Behavioral Research	246
Reliability, True Associations, and Observed Associations	246
Measurement Error (Low Reliability) Attenuates the Observed Associations Between Measures	249
Reliability, Effect Sizes, and Statistical Significance	254
Implications for Conducting and Interpreting Behavioral Research	259
Summary	263
Test Construction and Refinement	263
Item Discrimination and Other Information Regarding Internal Consistency	265
Item Difficulty (Mean) and Item Variance	270
Technical Appendix: R Syntax	271
Summary	277
Suggested Readings	278
PART III • VALIDITY	279
Chapter 8 • Validity: Conceptual Basis	281
What Is Validity?	282
The Importance of Validity	287

Validity Evidence: Test Content	289
Expert Rating Evidence	290
Threats to Content Validity	292
Content Validity Versus Face Validity	293
Validity Evidence: Internal Structure of the Test	294
Factor-Analytic Evidence	296
Validity Evidence: Response Processes	299
Direct Evidence	302
Indirect Evidence	303
Validity Evidence: Associations With Other Variables	304
Convergent Evidence	306
Discriminant Evidence	307
Criterion, Concurrent, and Predictive Evidence	309
Validity Evidence: Consequences of Testing	310
Evidence of Intended Effects	312
Evidence Regarding Unintended Differential Impact on Groups	313
Evidence Regarding Unintended Systemic Effects	314
Other Perspectives on Validity	316
Contrasting Reliability and Validity	319
Summary	320
Suggested Readings	321
Chapter 9 • Estimating and Evaluating Convergent and Discriminant Validity Evidence	323
A Construct's Nomological Network	324
Methods for Evaluating Convergent and Discriminant Validity	326
Focused Associations	327
Sets of Correlations	330
Multitrait–Multimethod Matrices	333
Quantifying Construct Validity	341
Factors Affecting a Validity Coefficient	347
Associations Between Constructs	348
Random Measurement Error and Reliability	348
Restricted Range	350
Skew and Relative Proportions	356
Method Variance	361
Time	362
Predictions of Single Events	362
Interpreting a Validity Coefficient	364
Squared Correlations and "Variance Explained"	364
Estimating Practical Effects: Binomial Effect Size Display, Taylor-Russell Tables, Utility Analysis, and Sensitivity/Specificity	367
Guidelines or Norms for a Field	376
Statistical Significance	378
Technical Appendix: R Syntax	385
Summary	390

Suggested Readings	391
Notes	392
PART IV • THREATS TO PSYCHOMETRIC QUALITY	393
Chapter 10 • Response Biases	395
Types of Response Biases	396
Acquiescence Bias ("Yea-Saying and Nay-Saying")	396
Extreme and Moderate Responding	404
Social Desirability ("Faking Good")	408
Malingering ("Faking Bad")	415
Careless or Random Responding	416
Guessing	420
Methods for Coping With Response Biases	421
Minimizing the Existence of Bias by Managing the Testing Context	421
Minimizing the Existence of Bias by Managing Test Content	423
Minimizing the Effects of Bias by Managing Test Content or Scoring	426
Managing Test Content to Detect Bias and Intervene	432
Using Specialized Tests to Detect Bias and Intervene	435
Response Biases, Response Sets, and Response Styles	437
Summary	437
Suggested Readings	438
Chapter 11 • Test Bias	441
Why Worry About Test Score Bias?	443
Detecting Construct Bias: Internal Evaluation of a Test	444
Reliability	446
Rank Order	447
Item Discrimination Index	448
Factor Analysis	450
Differential Item Functioning Analyses	452
Summary	456
Detecting Predictive Bias: External Evaluation of a Test	456
Basics of Regression Analysis	458
One Size Fits All: The Common Regression Equation	461
Intercept Bias	462
Slope Bias	466
Intercept and Slope Bias	470
Criterion Score Bias	470
The Effect of Reliability	471
Other Statistical Procedures	471
Test Fairness	472
Example: Is the SAT Biased in Terms of Race or Socioeconomic Status?	473
Race/Ethnicity	473
Socioeconomic Status	475
Technical Appendix: R Syntax	479
Summary	487

Suggested Readings	487
Notes	488
PART V • ADVANCED PSYCHOMETRIC APPROACHES	489
Chapter 12 • Confirmatory Factor Analysis	491
On the Use of EFA and CFA	493
The Frequency and Roles of EFA and CFA	493
Using CFA to Evaluate Measurement Models	493
The Process of CFA for Analysis of a Scale's Internal Structure	494
Overview of CFA and an Example	494
Preliminary Steps	496
Step 1: Specification of the Measurement Model	497
Step 2: Computations	500
Step 3: Interpreting and Reporting Output	503
Step 4: Model Modification and Reanalysis (If Necessary)	508
Comparing Models	510
Summary	511
CFA and Reliability	511
Evaluating Types of Classical Test Theory Measurement Models	511
Estimating Reliability (Omega Index)	515
CFA and Validity	518
CFA and Measurement Invariance	520
The Meaning of Measurement Invariance	521
Levels of Invariance: Meaning and Detection	523
Technical Appendix: R Syntax	531
Summary	542
Suggested Readings	543
Chapter 13 • Generalizability Theory	545
Multiple Facets of Measurement	547
Generalizability, Universes, and Variance Components	550
G Studies and D Studies	553
Conducting and Interpreting Generalizability Theory	
Analysis: A One-Facet Design	554
Phase 1: G Study	556
Phase 2: D Study	559
Conducting and Interpreting Generalizability Theory	
Analysis: A Two-Facet Design	563
Phase 1: G Study	565
Phase 2: D Study	571
Other Measurement Designs	573
Number of Facets	573
Random Versus Fixed Facets	574
Crossed Versus Nested Designs	577
Relative Versus Absolute Decisions	577

A Practical, Consistency-Oriented Interpretation of Variance Components	580
Systematic Variance Components Reflect "Consistent Variance"	580
Residual/Error Variance Component Reflects Inconsistent Variance	581
Generalizability Coefficients as the Proportion of Variance That Is Consistent	582
More Complex Designs	583
Technical Appendix: R Syntax	583
Summary	591
Suggested Readings	592
Notes	592
Chapter 14 • Item Response Theory and Rasch Models	593
Factors Affecting Responses to Test Items	593
Respondent Trait Level as a Determinant of Item Responses	594
Item Difficulty as a Determinant of Item Responses	595
Item Discrimination as a Determinant of Item Responses	597
Guessing	598
IRT Measurement Models	599
One-Parameter Logistic Model (or Rasch Model)	600
Two-Parameter Logistic Model	602
Three-Parameter Logistic Model	604
Graded Response Model	605
Obtaining Parameter Estimates: A 1PL Example	609
Model Fit	614
Item and Test Information	616
Item Characteristic Curves	616
Item Information and Test Information	619
Applications of IRT	626
Test Development and Improvement	626
Differential Item Functioning	626
Person Fit	627
Computerized Adaptive Testing	628
Technical Appendix: R Syntax	629
Summary	638
Suggested Readings	638
Glossary	641
References	649
Index	669