

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	xvii
Author's Acknowledgments	xx
Publisher's Acknowledgments	xxi
About the Author	xxiii
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	6
What Is a Psychological Test?	6
Types of Tests	7
Psychometrics	9
What Is Psychometrics?	9
A Brief History of Psychometrics	10
Challenges to Measurement in Psychology	11
The Importance of Individual Differences	15
But Psychometrics Goes Well Beyond "Differential" Psychology	16
Suggested Readings	17
PART I. BASIC CONCEPTS IN MEASUREMENT	19
Chapter 2. Scaling	21
Fundamental Issues With Numbers	22
The Property of Identity	22
The Property of Order	23

The Property of Quantity	24
The Number 0	25
Units of Measurement	27
Additivity and Counting	29
Additivity	29
Counts: When Do They Qualify as Measurement?	31
Four Scales of Measurement	31
Nominal Scales	32
Ordinal Scales	33
Interval Scales	33
Ratio Scales	34
Scales of Measurement: Practical Implications	35
Additional Issues Regarding Scales of Measurement	36
Summary	37
Suggested Readings	37
Chapter 3. Individual Differences and Correlations	39
The Nature of Variability	39
Importance of Individual Differences	40
Variability and Distributions of Scores	42
Central Tendency	43
Variability	43
Distribution Shapes and Normal Distributions	47
Quantifying the Association Between Distributions	49
Interpreting the Association Between Two Variables	49
Covariance	50
Correlation	53
Variance and Covariance for “Composite Variables”	54
Binary Items	55
Interpreting Test Scores	58
z Scores (Standard Scores)	60
Converted Standard Scores (Standardized Scores)	63
Percentile Ranks	64
Normalized Scores	67
Test Norms	68
Representativeness of the Reference Sample	69
Summary	70
Suggested Readings	71
Chapter 4. Test Dimensionality and Factor Analysis	73
Test Dimensionality	75
Three Dimensionality Questions	76
Unidimensional Tests	76
Multidimensional Tests With Correlated Dimensions (Tests With Higher-Order Factors)	78
Multidimensional Tests With Uncorrelated Dimensions	80
The Psychological Meaning of Test Dimensions	81

Factor Analysis: Examining the Dimensionality of a Test	82
The Logic and Purpose of Exploratory Factor Analysis:	
A Conceptual Overview	82
Conducting and Interpreting an Exploratory Factor Analysis	85
A Deeper Perspective on Factors, Factor Loadings, and Rotation	99
Factor Analysis of Binary Items	105
A Quick Look at Confirmatory Factor Analysis	105
Summary	106
Suggested Readings	107

PART II. RELIABILITY **109**

Chapter 5. Reliability: Conceptual Basis **111**

Overview of Reliability and Classical Test Theory	112
Observed Scores, True Scores, and Measurement Error	114
Variances in Observed Scores, True Scores, and Error Scores	117
Four Ways to Think of Reliability	119
Reliability as the Ratio of True Score Variance	
to Observed Score Variance	120
Reliability as Lack of Error Variance	122
Reliability as the (Squared) Correlation Between	
Observed Scores and True Scores	124
Reliability as the Lack of (Squared) Correlation	
Between Observed Scores and Error Scores	125
Reliability and the Standard Error of Measurement	127
From Theory to Practice: Measurement Models	
and Their Implications for Estimating Reliability	129
Overview of Key Assumptions	130
Parallel Tests	132
Tau-Equivalent and Essentially Tau-Equivalent Tests	136
Congeneric Tests	138
Tests With Correlated Errors	139
Summary	140
Domain Sampling Theory	140
Summary	141
Suggested Readings	142

Chapter 6. Empirical Estimates of Reliability **143**

Alternate Forms Reliability	144
Test–Retest Reliability	147
Internal Consistency Reliability	150
Split-Half Estimates of Reliability	151
"Raw" Coefficient Alpha	155
"Standardized" Coefficient Alpha	159
Raw Alpha for Binary Items: KR_{20}	162
Omega	163

On the Assumptions Underlying Alpha and Omega and the Relative Applicability of Those Indices	164
Internal Consistency Versus Dimensionality	167
Factors Affecting the Reliability of Test Scores	167
Sample Homogeneity and Reliability Generalization	174
Reliability of Difference Scores	175
Defining Difference Scores	176
Estimating the Reliability of Difference Scores	177
Factors Affecting the Reliability of Difference Scores	178
The Problem of Unequal Variability	180
Difference Scores: Summary and Caution	183
Summary	185
Note	186
Suggested Readings	186
Chapter 7. The Importance of Reliability	187
Applied Behavioral Practice: Evaluation of an Individual's Test Score	187
Point Estimates of True Scores	188
Confidence Intervals	191
Debate and Alternatives	193
Summary	194
Behavioral Research	194
Reliability, True Associations, and Observed Associations	195
Measurement Error (Low Reliability) Attenuates the Observed Associations Between Measures	197
Reliability, Effect Sizes, and Statistical Significance	201
Implications for Conducting and Interpreting Behavioral Research	205
Test Construction and Refinement	208
Item Discrimination and Other Information Regarding Internal Consistency	210
Item Difficulty (Mean) and Item Variance	214
Summary	215
Suggested Readings	216
PART III. VALIDITY	217
Chapter 8. Validity: Conceptual Basis	219
What Is Validity?	220
The Importance of Validity	224
Validity Evidence: Test Content	226
Expert Rating Evidence	227
Threats to Content Validity	228
Content Validity Versus Face Validity	229
Validity Evidence: Internal Structure of the Test	230
Factor-Analytic Evidence	231
Validity Evidence: Response Processes	234
Direct Evidence	236
Indirect Evidence	237

Validity Evidence: Associations With Other Variables	238
Convergent Evidence	239
Discriminant Evidence	240
Concurrent and Predictive Evidence	241
Validity Evidence: Consequences of Testing	242
Evidence of Intended Effects	244
Evidence Regarding Unintended Differential Impact on Groups	244
Evidence Regarding Unintended Systemic Effects	245
Other Perspectives on Validity	247
Contrasting Reliability and Validity	250
Summary	250
Suggested Readings	251

Chapter 9. Estimating and Evaluating Convergent and Discriminant Validity Evidence **253**

A Construct's Nomological Network	254
Methods for Evaluating Convergent and Discriminant Validity	256
Focused Associations	256
Sets of Correlations	259
Multitrait–Multimethod Matrices	262
Quantifying Construct Validity	269
Factors Affecting a Validity Coefficient	273
Associations Between Constructs	274
Random Measurement Error and Reliability	274
Restricted Range	276
Skew and Relative Proportions	281
Method Variance	286
Time	287
Predictions of Single Events	287
Interpreting a Validity Coefficient	288
Squared Correlations and "Variance Explained"	289
Estimating Practical Effects: Binomial Effect Size Display, Taylor-Russell Tables, Utility Analysis, and Sensitivity/Specificity	291
Guidelines or Norms for a Field	299
Statistical Significance	300
Summary	305
Notes	306
Suggested Readings	306

PART IV. THREATS TO PSYCHOMETRIC QUALITY **309**

Chapter 10. Response Biases **311**

Types of Response Biases	312
Acquiescence Bias ("Yea-Saying and Nay-Saying")	312
Extreme and Moderate Responding	317
Social Desirability ("Faking Good")	321
Malingering ("Faking Bad")	327

Careless or Random Responding	328
Guessing	330
Methods for Coping With Response Biases	331
Minimizing the Existence of Bias by Managing the Testing Context	332
Minimizing the Existence of Bias by Managing Test Content	333
Minimizing the Effects of Bias by Managing Test Content or Scoring	334
Managing Test Content to Detect Bias and Intervene	339
Using Specialized Tests to Detect Bias and Intervene	342
Response Biases, Response Sets, and Response Styles	344
Summary	344
Suggested Readings	344
Chapter 11. Test Bias	347
Why Worry About Test Score Bias?	349
Detecting Construct Bias: Internal Evaluation of a Test	349
Reliability	351
Rank Order	351
Item Discrimination Index	352
Factor Analysis	354
Differential Item Functioning Analyses	355
Summary	359
Detecting Predictive Bias: External Evaluation of a Test	359
Basics of Regression Analysis	361
One Size Fits All: The Common Regression Equation	363
Intercept Bias	364
Slope Bias	368
Intercept and Slope Bias	371
Criterion Score Bias	372
The Effect of Reliability	372
Other Statistical Procedures	372
Test Fairness	373
Example: Is the SAT Biased in Terms of Race or Socioeconomic Status?	374
Race/Ethnicity	374
Socioeconomic Status	376
Summary	379
Suggested Readings	380

PART V. ADVANCED PSYCHOMETRIC APPROACHES **383**

Chapter 12. Confirmatory Factor Analysis	385
On the Use of EFA and CFA	386
The Frequency and Roles of EFA and CFA	386
Using CFA to Evaluate Measurement Models	387

The Process of CFA for Analysis of a Scale's Internal Structure	388
Overview of CFA and an Example	388
Preliminary Steps	390
Step 1: Specification of the Measurement Model	390
Step 2: Computations	393
Step 3: Interpreting and Reporting Output	396
Step 4: Model Modification and Reanalysis (If Necessary)	401
Comparing Models	402
Summary	403
CFA and Reliability	403
Evaluating Types of CTT Measurement Models	403
Estimating Reliability (Omega Index)	406
CFA and Validity	410
CFA and Measurement Invariance	412
The Meaning of Measurement Invariance	412
Levels of Invariance: Meaning and Detection	414
Summary	421
Suggested Readings	421
Chapter 13. Generalizability Theory	423
Multiple Facets of Measurement	424
Generalizability, Universes, and Variance Components	427
G Studies and D Studies	429
Conducting and Interpreting Generalizability Theory Analysis:	
A One-Facet Design	430
Phase 1: G Study	431
Phase 2: D Study	434
Conducting and Interpreting Generalizability Theory Analysis:	
A Two-Facet Design	437
Phase 1: G Study	440
Phase 2: D Study	445
Other Measurement Designs	447
Number of Facets	447
Random Versus Fixed Facets	448
Crossed Versus Nested Designs	450
Relative Versus Absolute Decisions	451
Summary	453
Suggested Readings	454
Chapter 14. Item Response Theory and Rasch Models	455
Factors Affecting Responses to Test Items	455
Respondent Trait Level as a Determinant of Item Responses	456
Item Difficulty as a Determinant of Item Responses	456
Item Discrimination as a Determinant of Item Responses	458
Guessing	459
IRT Measurement Models	459
One-Parameter Logistic Model (or Rasch Model)	460

Two-Parameter Logistic Model	462
Three-Parameter Logistic Model	463
Graded Response Model	465
Obtaining Parameter Estimates: A 1PL Example	468
Model Fit	472
Item and Test Information	474
Item Characteristic Curves	474
Item Information and Test Information	477
Applications of IRT	483
Test Development and Improvement	483
Differential Item Functioning	484
Person Fit	484
Computerized Adaptive Testing	485
Summary	487
Suggested Readings	487
Glossary	489
References	501
Subject Index	523
Author Index	537