

# Contents

<b>1 • <u>Introduction</u></b>	<b>1</b>
1.1 Psychological Measurement and Tests	1
1.2 Tests and Samples of Behavior	3
1.3 Types of Tests	3
1.4 Origin of Psychometrics	4
1.5 Definition of Measurement	5
1.6 Measuring Behavior	5
1.7 Psychometrics and Its Importance to Research and Practice	7
1.8 Organization of This Book	9
Key Terms and Definitions	10
<b>2 • <u>Measurement and Statistical Concepts</u></b>	<b>13</b>
2.1 Introduction	13
2.2 Numbers and Measurement	13
2.3 Properties of Measurement in Relation to Numbers	14
2.4 Levels of Measurement	20
2.5 Contemporary View on the Levels of Measurement and Scaling	22
2.6 Statistical Foundations for Psychometrics	22
2.7 Variables, Frequency Distributions, and Scores	23
2.8 Summation or Sigma Notation	29
2.9 Shape, Central Tendency, and Variability of Score Distributions	31
2.10 Correlation, Covariance, and Regression	42
2.11 Summary	55
Key Terms and Definitions	55

**3 • Criterion, Content, and Construct Validity**

59

- 3.1 Introduction 59
- 3.2 Criterion Validity 63
- 3.3 Essential Elements of a High-Quality Criterion 64
- 3.4 Statistical Estimation of Criterion Validity 66
- 3.5 Correction for Attenuation 68
- 3.6 Limitations to Using the Correction for Attenuation 70
- 3.7 Estimating Criterion Validity with Multiple Predictors:  
Partial Correlation 70
- 3.8 Estimating Criterion Validity with Multiple Predictors:  
Higher-Order Partial Correlation 77
- 3.9 Coefficient of Multiple Determination and Multiple Correlation 80
- 3.10 Estimating Criterion Validity with More Than One Predictor:  
Multiple Linear Regression 84
- 3.11 Regression Analysis for Estimating Criterion Validity:  
Development of the Regression Equation 85
- 3.12 Unstandardized Regression Equation for Multiple Regression 87
- 3.13 Testing the Regression Equation for Significance 87
- 3.14 Partial Regression Slopes 90
- 3.15 Standardized Regression Equation 93
- 3.16 Predictive Accuracy of a Regression Analysis 94
- 3.17 Predictor Subset Selection in Regression 101
- 3.18 Summary 102  
Key Terms and Definitions 102

**4 • Statistical Aspects of the Validation Process**

105

- 4.1 Techniques for Classification and Selection 105
- 4.2 Discriminant Analysis 106
- 4.3 Multiple-Group Discriminant Analysis 114
- 4.4 Logistic Regression 117
- 4.5 Logistic Multiple Discriminant Analysis:  
Multinomial Logistic Regression 122
- 4.6 Model Fit in Logistic Regression 125
- 4.7 Content Validity 125
- 4.8 Limitations of the Content Validity Model 126
- 4.9 Construct Validity 126
- 4.10 Establishing Evidence of Construct Validity 127
- 4.11 Correlational Evidence of Construct Validity 130
- 4.12 Group Differentiation Studies of Construct Validity 131
- 4.13 Factor Analysis and Construct Validity 131
- 4.14 Multitrait–Multimethod Studies 134
- 4.15 Generalizability Theory and Construct Validity 136
- 4.16 Summary and Conclusions 137  
Key Terms and Definitions 138

<b>5 • <u>Scaling</u></b>	<b>141</b>
5.1 Introduction	141
5.2 A Brief History of Scaling	142
5.3 Psychophysical versus Psychological Scaling	144
5.4 Why Scaling Models Are Important	146
5.5 Types of Scaling Models	146
5.6 Stimulus-Centered Scaling	147
5.7 Thurstone's Law of Comparative Judgment	148
5.8 Response-Centered Scaling	150
5.9 Scaling Models Involving Order	150
5.10 Guttman Scaling	151
5.11 The Unfolding Technique	153
5.12 Subject-Centered Scaling	156
5.13 Data Organization and Missing Data	160
5.14 Incomplete and Missing Data	162
5.15 Summary and Conclusions	162
Key Terms and Definitions	162
<b>6 • <u>Test Development</u></b>	<b>165</b>
6.1 Introduction	165
6.2 Guidelines for Test and Instrument Development	166
6.3 Item Analysis	182
6.4 Item Difficulty	182
6.5 Item Discrimination	184
6.6 Point-Biserial Correlation	186
6.7 Biserial Correlation	188
6.8 Phi Coefficient	189
6.9 Tetrachoric Correlation	190
6.10 Item Reliability and Validity	190
6.11 Standard Setting	193
6.12 Standard-Setting Approaches	194
6.13 The Nedelsky Method	195
6.14 The Ebel Method	196
6.15 The Angoff Method and Modifications	196
6.16 The Bookmark Method	198
6.17 Summary and Conclusions	199
Key Terms and Definitions	199
<b>7 • <u>Reliability</u></b>	<b>203</b>
7.1 Introduction	203
7.2 Conceptual Overview	204
7.3 The True Score Model	206
7.4 Probability Theory, True Score Model, and Random Variables	207

- 7.5 Properties and Assumptions of the True Score Model 209
- 7.6 True Score Equivalence, Essential True Score Equivalence, and Congeneric Tests 219
- 7.7 Relationship between Observed and True Scores 219
- 7.8 The Reliability Index and Its Relationship to the Reliability Coefficient 221
- 7.9 Summarizing the Ways to Conceptualize Reliability 221
- 7.10 Reliability of a Composite 223
- 7.11 Coefficient of Reliability: Methods of Estimation Based on Two Occasions 228
- 7.12 Methods Based on a Single Testing Occasion 230
- 7.13 Estimating Coefficient Alpha: Computer Program and Example Data 234
- 7.14 Reliability of Composite Scores Based on Coefficient Alpha 238
- 7.15 Reliability Estimation Using the Analysis of Variance Method 240
- 7.16 Reliability of Difference Scores 241
- 7.17 Application of the Reliability of Difference Scores 243
- 7.18 Errors of Measurement and Confidence Intervals 244
- 7.19 Standard Error of Measurement 244
- 7.20 Standard Error of Prediction 250
- 7.21 Summarizing and Reporting Reliability Information 251
- 7.22 Summary and Conclusions 252  
Key Terms and Definitions 253

## **8 • Generalizability Theory**

257

- 8.1 Introduction 257
- 8.2 Purpose of Generalizability Theory 258
- 8.3 Facets of Measurement and Universe Scores 259
- 8.4 How Generalizability Theory Extends Classical Test Theory 260
- 8.5 Generalizability Theory and Analysis of Variance 260
- 8.6 General Steps in Conducting a Generalizability Theory Analysis 263
- 8.7 Statistical Model for Generalizability Theory 263
- 8.8 Design 1: Single-Facet Person-by-Item Analysis 266
- 8.9 Proportion of Variance for the  $p \times i$  Design 271
- 8.10 Generalizability Coefficient and CTT Reliability 273
- 8.11 Design 2: Single-Facet Crossed Design with Multiple Raters 274
- 8.12 Design 3: Single-Facet Design with the Same Raters on Multiple Occasions 278
- 8.13 Design 4: Single-Facet Nested Design with Multiple Raters 279
- 8.14 Design 5: Single-Facet Design with Multiple Raters Rating on Two Occasions 280
- 8.15 Standard Errors of Measurement: Designs 1–5 281
- 8.16 Two-Facet Designs 281
- 8.17 Summary and Conclusions 286  
Key Terms and Definitions 287

<b>9 • <u>Factor Analysis</u></b>	<b>289</b>
9.1 Introduction	289
9.2 Brief History	291
9.3 Applied Example with GfGc Data	292
9.4 Estimating Factors and Factor Loadings	294
9.5 Factor Rotation	301
9.6 Correlated Factors and Simple Structure	306
9.7 The Factor Analysis Model, Communality, and Uniqueness	309
9.8 Components, Eigenvalues, and Eigenvectors	312
9.9 Distinction between Principal Components Analysis and Factor Analysis	315
9.10 Confirmatory Factor Analysis	319
9.11 Confirmatory Factor Analysis and Structural Equation Modeling	319
9.12 Conducting Factor Analysis: Common Errors to Avoid	322
9.13 Summary and Conclusions	325
Key Terms and Definitions	325
<b>10 • <u>Item Response Theory</u></b>	<b>329</b>
10.1 Introduction	329
10.2 How IRT Differs from CTT	330
10.3 Introduction to IRT	331
10.4 Strong True Score Theory, IRT, and CTT	332
10.5 Philosophical Views on IRT	333
10.6 Conceptual Explanation of How IRT Works	334
10.7 Assumptions of IRT Models	336
10.8 Test Dimensionality and IRT	337
10.9 Type of Correlation Matrix to Use in Dimensionality Analysis	337
10.10 Dimensionality Assessment Specific to IRT	341
10.11 Local Independence of Items	345
10.12 The Invariance Property	349
10.13 Estimating the Joint Probability of Item Responses Based on Ability	351
10.14 Item and Ability Information and the Standard Error of Ability	358
10.15 Item Parameter and Ability Estimation	362
10.16 When Traditional IRT Models Are Inappropriate to Use	364
10.17 The Rasch Model	366
10.18 The Rasch Model, Linear Models, and Logistic Regression Models	366
10.19 Properties and Results of a Rasch Analysis	371
10.20 Item Information for the Rasch Model	373
10.21 Data Layout	373
10.22 One-Parameter Logistic Model for Dichotomous Item Responses	374
10.23 Two-Parameter Logistic Model for Dichotomous Item Responses	381
10.24 Item Information for the Two-Parameter Model	388
10.25 Three-Parameter Logistic Model for Dichotomous Item Responses	389
10.26 Item Information for the Three-Parameter Model	397

10.27	Choosing a Model: A Model Comparison Approach	400
10.28	Summary and Conclusions	404
	Key Terms and Definitions	404
<b>11 •</b>	<b><u>Norms and Test Equating</u></b>	<b>407</b>
11.1	Introduction	407
11.2	Norms, Norming, and Norm-Referenced Testing	408
11.3	Planning a Norming Study	408
11.4	Scaling and Scale Scores	410
11.5	Standard Scores under Linear Transformation	411
11.6	Percentile Rank Scale	415
11.7	Interpreting Percentile Ranks	416
11.8	Normalized z- or Scale Scores	418
11.9	Common Standard Score Transformations or Conversions	422
11.10	Age- and Grade-Equivalent Scores	424
11.11	Test Score Linking and Equating	425
11.12	Techniques for Conducting Equating: Linear Methods	428
11.13	Design I: Random Groups—One Test Administered to Each Group	429
11.14	Design II: Random Groups with Both Tests Administered to Each Group, Counterbalanced (Equally Reliable Tests)	432
11.15	Design III: One Test Administered to Each Study Group, Anchor Test Administered to Both Groups (Equally Reliable Tests)	435
11.16	Equipercntile Equating	436
11.17	Test Equating Using IRT	439
11.18	IRT True Score Equating	443
11.19	Observed Score, True Score, and Ability	445
11.20	Summary and Conclusions	447
	Key Terms and Definitions	448
<b>Appendix •</b>	<b><u>Mathematical and Statistical Foundations</u></b>	<b>451</b>
	<b>References</b>	<b>519</b>
	<b>Author Index</b>	<b>531</b>
	<b>Subject Index</b>	<b>537</b>
	<b>About the Author</b>	<b>552</b>

The companion website [www.guilford.com/price2-materials](http://www.guilford.com/price2-materials)  
presents datasets for all examples as well as  
PowerPoints of figures and key concepts.