

# Contents

<b>About the Authors</b>	xv
<b>Preface</b>	xvii
<b>PART 1: Introduction</b>	<b>1</b>
Module 1. Introduction and Overview	3
What Makes Tests Useful	4
Individual Differences	6
Constituents in the Testing Process	7
Diversity Issues	8
Reasons for Concern	9
Concluding Comments	9
Best Practices	10
Practical Questions	10
Case Studies	10
Exercises	12
Further Readings	13
Module 2. Statistics Review for Psychological Measurement	15
Organizing and Displaying Test Data	15
Univariate Descriptive Statistics	19
Bivariate Descriptive Statistics	22
Estimation	25
Concluding Comments	26
Best Practices	26
Practical Questions	27
Case Studies	27
Exercises	29
Further Readings	30

Module 3. Psychological Scaling	31
Levels of Measurement	32
Unidimensional Scaling Models	33
Multidimensional Scaling Models	35
A Step-by-Step Scaling Example	36
Concluding Comments	38
Best Practices	38
Practical Questions	38
Case Studies	39
Exercises	41
Further Readings	42
 Module 4. Test Preparation and Specification	 43
Step 1: Specify the Type of Measure	43
Step 2: Define the Domain of the Test	44
Step 3: Determine Whether a Closed-Ended or Open-Ended Format Will Be Used	45
Step 4: Determine the Item Format	46
Step 5: Determine Whether the Exam Will Be Administered Individually or in a Group Setting	46
Step 6: Determine the Appropriate Test Length	46
Step 7: Determine the Appropriate Difficulty Level of Items That Will Compose the Test	47
Concluding Comments	47
Best Practices	47
Practical Questions	48
Case Studies	48
Exercises	51
Further Readings	53
 <b>PART 2: Reliability, Validity, and Test Bias</b>	 <b>55</b>
 Module 5. Reliability Overview: Classical Test Theory	 57
Interpreting the Size of the Reliability Coefficient	59
Reliability and Test Length	60
Concluding Comments	61
Best Practices	61
Practical Questions	61
Case Studies	62
Exercises	66
Further Readings	67

Module 6. Estimating Reliability	69
Sources of Measurement Error	69
Types of Reliability	70
What Do We Do With the Reliability Estimates Now That We Have Them?	74
Concluding Comments	75
Best Practices	76
Practical Questions	76
Case Studies	77
Exercises	80
Further Readings	81
Module 7. Content Validity	83
An Important Note	83
The Role of Expert Judgment	83
Content Validity: A Simple Example	84
Formalizing Content Validity With the Content Validity Ratio	85
Acceptable Size of the CVR	85
A CVR Computational Example	86
The Content Validity Index	86
A Second Approach to Formalizing Content Validity	87
Face Validity	87
Concluding Comments	87
Best Practices	88
Practical Questions	88
Case Studies	89
Exercises	91
The Undergraduate Academic Work Ethic Measure	93
Further Readings	94
Module 8. Criterion-Related Validation	95
Criterion-Related Validation Research Designs	95
Examples of Criterion-Related Validation	96
Interpreting the Validity Coefficient	97
Attenuation and Inflation of Observed Validity Coefficients	98
Additional Considerations	102
Concluding Comments	102
Best Practices	102
Practical Questions	103
Case Studies	104
Exercises	107
Further Readings	110

Module 9. Construct Validity	111
Multitrait-Multimethod Matrices	112
Additional Aspects of Construct Validation	114
A Contemporary Conception of Construct Validation	115
Concluding Comments	117
Best Practices	117
Practical Questions	117
Case Studies	118
Exercises	121
Further Readings	122
 Module 10. Validity Generalization and Psychometric Meta-Analysis	 123
Validity Generalization	124
From VG to Psychometric Meta-Analysis	125
Conducting a Meta-Analysis	125
A Step-by-Step Meta-Analysis Example	127
Concluding Comments	130
Best Practices	130
Practical Questions	131
Case Studies	131
Exercises	134
Further Readings	136
 Module 11. Test Bias, Unfairness, and Equivalence	 137
Establishing Test Bias	137
Test Fairness	139
A Step-by-Step Example of Estimating Test Bias	141
Test Equivalence	144
Testing Individuals with Disabilities	145
Levels of Accommodation	146
Concluding Comments	146
Best Practices	147
Practical Questions	147
Case Studies	148
Exercises	151
Further Readings	152
 <b>PART 3: Practical Issues in Test Construction</b>	 <b>153</b>
 Module 12. Developing Tests of Maximal Performance	 155
Getting Started	155
Speed Versus Power Tests	156

Level of Cognitive Objective	156
Item Format	158
Item Writing	159
The Process of Item Modification	162
Concluding Comments	162
Best Practices	163
Practical Questions	163
Case Studies	163
Exercises	168
Further Readings	170
Module 13. Classical Test Theory Item Analysis	171
Item Difficulty	171
Item Discrimination	172
Norm-Referenced Versus Criterion-Referenced Tests	175
Overall Test Statistics	176
A Step-by-Step Example of an Item Analysis	176
Concluding Comments	182
Best Practices	182
Practical Questions	183
Case Studies	183
Exercises	186
Further Readings	193
Module 14. Scoring Tests	195
Judgmental Methods	195
Judgmental/Empirical Methods	197
Empirical/Judgmental Methods	198
Subgroup Norming and Banding	199
A Step-by-Step Example of Setting Cutoff Scores	200
Concluding Comments	202
Best Practices	203
Practical Questions	203
Case Studies	204
Exercises	206
Further Readings	216
Module 15. Developing Measures of Typical Performance	217
A Necessary Forewarning	217
Test Specifications	218
Item Writing	221
Rational or Empirical Test Development?	222

Pilot Testing	223
Survey Implementation	224
Concluding Comments	224
Best Practices	225
Practical Questions	225
Case Studies	226
Exercises	229
Further Readings	233
 Module 16. Response Biases	 235
Guessing on Knowledge and Aptitude Tests	235
Correcting for Guessing on Knowledge Tests	236
Response Biases on Personality and Attitude Measures	238
A Step-by-Step Example of Identifying and Examining Response Biases	240
Concluding Comments	242
Best Practices	242
Practical Questions	243
Case Studies	243
Exercises	245
Further Readings	247
 <b>PART 4: Advanced Topics</b>	 <b>249</b>
 Module 17. Combining Predictors Using Multiple Regression	 251
The Multiple Regression Equation	251
The Multiple Regression Equation: An Example	252
Prediction Accuracy	254
Predictor Interrelationships	255
Stability of the Validity Coefficient	257
Adequate Sample Size	259
Concluding Comments	260
Best Practices	260
Practical Questions	260
Case Studies	260
Exercises	264
Further Readings	266
 Module 18. Exploratory Factor Analysis	 269
Exploratory Factor Analysis	269
Interpreting EFA Output	273
A Step-by-Step Example of EFA	274

Concluding Comments	278
Best Practices	278
Practical Questions	278
Case Studies	279
Exercises	282
Further Readings	286
Module 19. Confirmatory Factor Analysis	287
Testing the Hypothesized Factor Model	289
A Step-by-Step Example of CFA	290
Revising Your Original Model	292
Item Bundling	293
Concluding Comments	294
Best Practices	294
Practical Questions	295
Case Studies	295
Exercises	299
Further Readings	300
Module 20. Item Response Theory	301
IRT Versus CTT	301
General Overview of IRT	302
Item Response Functions	304
Item Information Functions	307
A Step-by-Step Example of Conducting an IRT Analysis	309
Concluding Comments	315
Best Practices	315
Practical Questions	316
Case Studies	316
Exercises	319
Further Readings	323
Module 21. Applications of Item Response Theory:	
Computer Adaptive Testing and Differential Item Functioning	325
Computer Adaptive Testing	325
Differential Item Functioning	328
Concluding Comments	331
Best Practices	332
Practical Questions	332
Case Studies	332
Exercises	335
Further Readings	336

Module 22. Generalizability Theory	337
The GT Framework	338
Distinctions Within the GT Framework	341
Conducting a GT analysis	343
Concluding Comments	343
Best Practices	344
Practical Questions	344
Case Studies	345
Exercises	347
Further Readings	348
<b>Appendix A</b>	
<b>Course-Long Exercise on Psychological Scale Development</b>	349
<b>Appendix B</b>	
<b>Data Set Descriptions</b>	357
<b>Glossary of Key Terms</b>	367
<b>References</b>	379
<b>Author Index</b>	389
<b>Subject Index</b>	393