

Contents

- 1 Introduction to health measurement scales 1**
 - Introduction to measurement 1
 - A roadmap to the book 3
- 2 Basic concepts 7**
 - Introduction to basic concepts 7
 - Searching the literature 7
 - Critical review 8
 - Empirical forms of validity 10
 - The two traditions of assessment 14
 - Summary 17
- 3 Devising the items 19**
 - Introduction to devising items 19
 - The source of items 20
 - Content validity 25
 - Generic versus specific scales and the 'fidelity versus bandwidth' issue 28
 - Translation 30
- 4 Scaling responses 38**
 - Introduction to scaling responses 38
 - Some basic concepts 38
 - Categorical judgements 39
 - Continuous judgements 41
 - To rate or to rank 64
 - Multidimensional scaling 65
- 5 Selecting the items 74**
 - Introduction to selecting items 74
 - Interpretability 74
 - Face validity 79
 - Frequency of endorsement and discrimination 80
 - Homogeneity of the items 81
 - Multifactor inventories 91
 - When homogeneity does not matter 92
 - Putting it all together 94

- 6 Biases in responding 100**
 - Introduction to biases in responding 100
 - The differing perspectives 100
 - Answering questions: the cognitive requirements 101
 - Optimizing and satisficing 104
 - Social desirability and faking good 106
 - Deviation and faking bad 111
 - Minimizing biased responding 111
 - Yea-saying or acquiescence 115
 - End-aversion, positive skew, and halo 115
 - Framing 118
 - Biases related to the measurement of change 119
 - Reconciling the two positions 121
 - Proxy reporting 121
 - Testing the items 122
- 7 From items to scales 131**
 - Introduction to from items to scales 131
 - Weighting the items 131
 - Missing items 134
 - Multiplicative composite scores 135
 - Transforming the final score 138
 - Age and sex norms 143
 - Establishing cut points 145
 - Receiver operating characteristic curves 149
 - Summary 156
- 8 Reliability 159**
 - Introduction to reliability 159
 - Basic concepts 159
 - Philosophical implications 161
 - Terminology 164
 - Defining reliability 164
 - Other considerations in calculating the reliability of a test: measuring consistency or absolute agreement 167
 - The observer nested within subject 169
 - Multiple observations 170
 - Other types of reliability 171
 - Different forms of the reliability coefficient 172
 - Kappa coefficient versus the ICC 178

- The method of Bland and Altman 179
- Issues of interpretation 180
- Improving reliability 185
- Standard error of the reliability coefficient and sample size 187
- Reliability generalization 192
- Summary 196
- 9 Generalizability theory 200**
 - Introduction to generalizability theory 200
 - Generalizability theory fundamentals 202
 - An example 204
 - The first step—the ANOVA 204
 - Step 2—from ANOVA to G coefficients 207
 - Step 3—from G study to D study 212
 - ANOVA for statisticians and ANOVA for psychometricians 212
 - Confidence intervals for G coefficients 214
 - Getting the computer to do it for you 214
 - Some common examples 215
 - Uses and abuses of G theory 224
 - Summary 225
- 10 Validity 227**
 - Introduction to validity 227
 - Why assess validity? 227
 - Reliability and validity 228
 - A history of the ‘types’ of validity 229
 - Content validation 232
 - Criterion validation 233
 - Construct validation 235
 - Responsiveness and sensitivity to change 244
 - Validity and ‘types of indices’ 244
 - Biases in validity assessment 245
 - Validity generalization 250
 - Summary 250
- 11 Measuring change 254**
 - Introduction to measuring change 254
 - The goal of measurement of change 254
 - Why not measure change directly? 255
 - Measures of association—reliability and sensitivity to change 256
 - Difficulties with change scores in experimental designs 261

Change scores and quasi-experimental designs 262
 Measuring change using multiple observations: growth curves 264
 How much change is enough? 268
 Summary 269

12 Item response theory 273

Introduction to item response theory 273
 Problems with classical test theory 273
 The introduction of item response theory 275
 A note about terminology 275
 Item calibration 276
 The one-parameter model 280
 The two- and three-parameter models 282
 Polytomous models 284
 Item information 286
 Item fit 287
 Person fit 289
 Differential item functioning 289
 Unidimensionality and local independence 290
 Test information and the standard error of measurement 294
 Equating tests 295
 Sample size 296
 Mokken scaling 296
 Advantages 297
 Disadvantages 299
 Computer programs 300

13 Methods of administration 304

Introduction to methods of administration 304
 Face-to-face interviews 304
 Telephone questionnaires 307
 Mailed questionnaires 312
 The necessity of persistence 317
 Computerized administration 319
 Using e-mail and the Web 322
 Personal data assistants and smart phones 328
 From administration to content: the impact of technology on scale construction 329
 Reporting response rates 331

14 Ethical considerations 340

Introduction to ethical considerations 340

Informed consent	341
Freedom of consent	344
Confidentiality	345
Consequential validation	346
Summary	347
15 Reporting test results	349
Introduction to reporting test results	349
Standards for educational and psychological testing	350
The STARD initiative	352
GRRAS	354
Summary	354
Appendices	
Appendix A Where to find tests	357
Appendix B A (very) brief introduction to factor analysis	375
Author Index	381
Subject Index	391