Contents

Acknowledgements vii

List of figures ix

List of tables xi

Preface xiii

1 Testing and assessment in context 1
1. Test purpose 1
2. Tests in educational systems 4
3. Testing rituals 5
4. Unintended consequences 6
5. Testing and society 8
6. Historical interlude I 11
7. The politics of language testing 12
8. Historical interlude II 15
9. Professionalising language education and testing 17
10. Validity 19

Activities 21

2 Standardised testing 31
1. Two paradigms 31
2. Testing as science 32
3. What's in a curve? 35
4. The curve and score meaning 36
5. Putting it into practice 37
6. Test scores in a consumer age 42
7. Testing the test 44
8. Introducing reliability 46
9. Calculating reliability 47
10. Living with uncertainty 54
11. Reliability and test length 57
12. Relationships with other measures 57
13. Measurement 59

Activities 60

3 Classroom assessment 67
1. Life at the chalk-face 67
2. Assessment for Learning 68
3. Self- and peer-assessment 70
4. Dynamic Assessment 72
5. Understanding change 75
6. Assessment and second language acquisition 77
7. Criterion-referenced testing 79
8. Dependability 81
9. Some thoughts on theory 87
Activities 90

4 Deciding what to test 93
1. The test design cycle 93
2. Construct definition 96
3. Where do constructs come from? 102
4. Models of communicative competence 105
5. From definition to design 118
Activities 120

5 Designing test specifications 127
1. What are test specifications? 127
2. Specifications for testing and teaching 134
3. A sample detailed specification for a reading test 139
4. Granularity 147
5. Performance conditions 148
6. Target language use domain analysis 149
7. Moving back and forth 154
Activities 155

6 Evaluating, prototyping and piloting 159
1. Investigating usefulness and usability 159
2. Evaluating items, tasks and specifications 159
3. Guidelines for multiple-choice items 172
4. Prototyping 173
5. Piloting 179
6. Field testing 185
7. Item shells 186
8. Operational item review and pre-testing 188
Activities 190

7 Scoring language tests 197
1. Scoring items 197

2. Scorability 201
3. Scoring constructed response tasks 208
4. Automated scoring 216
5. Corrections for guessing 218
6. Avoiding own goals 219
Activities 220

8 Aligning tests to standards 225
1. It's as old as the hills 225
2. The definition of 'standards' 225
3. The uses of standards 226
4. Unintended consequences revisited 228
5. Using standards for harmonisation and identity 229
6. How many standards can we afford? 231
7. Performance level descriptors (PLDs) and test scores 233
8. Some initial decisions 234
9. Standard-setting methodologies 236
10. Evaluating standard setting 241
11. Training 243
12. The special case of the CEFR 244
13. You can always count on uncertainty 248
Activities 250

9 Test administration 253
1. No, no. Not me! 253
2. Controlling extraneous variables 254
3. Rituals revisited 258
4. Standardised conditions and training 259
5. Planned variation: accommodations 262
6. Unplanned variation: cheating 264
7. Scoring and moderation 267
8. Data handling and policy 268
9. Reporting outcomes to stakeholders 269
10. The expense of it all 272
Activities 274

10 Testing and teaching 277
1. The things we do for tests 277
2. Washback 277
3. Washback and content alignment 282
4. Preparing learners for tests 288
5. Selecting and using tests 292
6. The gold standard 295

Activities 298

Epilogue 300

Appendices 301

Glossary 319

References 325

Index 343