## Contents

Acknowledgements	
List of figures	ix
List of tables	xi
Preface	xiii
<ol> <li>Testing and assessment in context</li> <li>Test purpose</li> <li>Tests in educational systems</li> <li>Testing rituals</li> <li>Unintended consequences</li> <li>Testing and society</li> <li>Historical interlude I</li> <li>The politics of language testing</li> <li>Historical interlude II</li> <li>Professionalising language education and testing</li> <li>Validity</li> <li>Activities</li> </ol>	1 1 4 5 6 8 11 12 15 17 19 21
<ol> <li>Standardised testing</li> <li>Two paradigms</li> <li>Testing as science</li> <li>What's in a curve?</li> <li>The curve and score meaning</li> <li>Putting it into practice</li> <li>Test scores in a consumer age</li> <li>Testing the test</li> <li>Introducing reliability</li> </ol>	31 31 32 35 36 37 42 44
<ul> <li>9. Calculating reliability</li> <li>10. Living with uncertainty</li> <li>11. Reliability and test length</li> <li>12. Relationships with other measures</li> <li>13. Measurement</li> <li>Activities</li> </ul>	47 54 57 57 59

## iv Contents

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
Ac <sup>1</sup>	tivities	90
41	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
Ac	tivities	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
Ac	tivities	158
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	188
7.	Item shells	186
8.	Operational item review and pre-testing	188
Ac	tivities	190
7	Scoring language tests	197
1.	Scoring items	197

-	75 37	
( or	itents	
		V

2. 3. 4. 5. 6. Acti	Scorability Scoring constructed response tasks Automated scoring Corrections for guessing Avoiding own goals vities	201 208 216 218 219 220
1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13.	It's as old as the hills The definition of 'standards' The uses of standards Unintended consequences revisited Using standards for harmonisation and identity How many standards can we afford? Performance level descriptors (PLDs) and test scores Some initial decisions Standard-setting methodologies Evaluating standard setting Training The special case of the CEFR You can always count on uncertainty vities	225 225 226 228 229 231 233 234 236 241 243 244 248 250
1. 2. 3. 4. 5. 6. 7. 8. 9.	No, no. Not me! Controlling extraneous variables Rituals revisited Standardised conditions and training Planned variation: accommodations Unplanned variation: cheating Scoring and moderation Data handling and policy Reporting outcomes to stakeholders The expense of it all vities	253 253 254 258 259 262 264 267 268 269 272 274
	Testing and teaching The things we do for tests Washback Washback and content alignment Preparing learners for tests Selecting and using tests The gold standard	277 277 277 282 288 292 295

## vi Contents

Activities	298
Epilogue	300
Appendices	301
Glossary	319
References	325
Index	3/19