Table of Contents

- Acknowledgements p. x
- Preface p. xi
- 1 Introduction: the rubber ruler p. 1
- 1.1 Why test language learning? p. 1
- 1.2 What is a language test? p. 2
- 1.2.1 What are the properties of measuring devices? p. 2
- 1.3 The rubber ruler p. 3
- 1.4 Tests, measurement and evaluation p. 5
- 1.4.1 Evaluation without measurement p. 5
- 1.4.2 Measurement without a test p. 5
- 1.4.3 Tests p. 6
- 1.5 Ethical uses of language tests p. 9
- 1.5.1 Reliability p. 10
- 1.5.2 Validity p. 10
- Further reading p. 13
- Exercises p. 13
- 2 Measuring language ability and making decisions p. 17
- 2.1 Measuring language ability p. 17
- 2.2 Making decisions about learners, teachers, programmes, and policies p. 19
- 2.3 Contexts of language use p. 20
- 2.3.1 Contextual features p. 21
- 2.3.2 Authenticity p. 24
- 2.4 Making valid interpretations of test performance p. 26
- 2.4.1 Consistency of measurement p. 26
- 2.4.2 Validity: evidence for interpretations of test performance p. 29
- 2.5 Conclusion: bias for best p. 34
- Further reading p. 35
- Exercises p. 35
- 3 Test development p. 38
- 3.1 What do I need to test? Needs analysis p. 38
- 3.1.1 Define the purpose of the test p. 40
- 3.1.2 Conduct a preliminary investigation p. 40
- 3.1.3 Collect primary data p. 40
- 3.1.4 Collect secondary data p. 42
- 3.1.5 Analyse target language use task and language characteristics p. 44
- 3.2 How am I going to test language abilities? Turning target language use tasks into test tasks p. 48
- 3.2.1 Developing a test task p. 48
- 3.2.2 Developing a blueprint for the test p. 49
- 3.2.3 Options for test tasks p. 49
- 3.3 How am I going to give the test? Test administration p. 54
- 3.3.1 Test environment p. 54
- 3.3.2 Personnel p. 54
- 3.3.3 Procedures p. 55

- 3.3.4 Scoring p. 55
- 3.4 How can my computer assist me in test development? Computer-based tools p. 56
- 3.4.1 Hot Potatoes p. 57
- 3.4.2 Moodle p. 60
- 3.4.3 WebCT p. 60
- 3.5 Conclusion p. 63
- Further reading p. 64
- Exercises p. 64
- 4 Alternatives in assessment p. 67
- 4.1 Norm-referenced and criterion-referenced tests p. 67
- 4.2 Communicative language tests p. 69
- 4.3 'General' and 'specific purpose' language tests p. 70
- 4.4 Discrete-point and integrative tests p. 70
- 4.5 Formative and summative assessment p. 72
- 4.6 Alternative approaches to assessment p. 73
- 4.6.1 Conference assessments p. 74
- 4.6.2 Portfolio assessment p. 74
- 4.6.3 Self- and peer-assessments p. 75
- 4.6.4 Task-based and performance assessment p. 76
- 4.6.5 Dynamic assessment p. 79
- 4.6.6 Summary p. 80
- 4.7 Conclusion p. 80
- Further reading p. 81
- Exercises p. 81
- 5 By the numbers: a statistics mini-course p. 85
- 5.1 Introduction p. 85
- 5.2 Normal distribution p. 87
- 5.3 The average or mean p. 88
- 5.4 Standard deviation p. 90
- 5.4.1 Standard deviation as a unit of measurement p. 92
- 5.5 Correlation p. 93
- 5.6 Probability and statistical significance p. 97
- 5.7 The t-test of the difference between two averages p. 99
- 5.8 Analysis of variance p. 101
- 5.9 Reliability p. 104
- 5.9.1 Split-half method p. 105
- 5.9.2 Internal consistency method p. 106
- 5.9.3 Standard error of measurement p. 108
- 5.10 The reliability of human raters p. 110
- 5.11 Conclusion p. 111
- Further reading p. 112
- Exercises p. 113
- 6 Technology and language testing p. 115
- 6.1 Introduction p. 115
- 6.2 Issues in technology and language testing p. 116
- 6.2.1 Technology and test taker attitudes p. 117

- 6.2.2 Language performance and different media p. 117
- 6.2.3 Technology and the construct to be measured p. 118
- 6.2.4 Technology and assessment tasks p. 118
- 6.2.5 The limits of automated scoring p. 119
- 6.3 Technology and language task types p. 119
- 6.3.1 Listening tasks p. 119
- 6.3.2 Integrated listening and speaking tasks p. 122
- 6.3.3 Writing tasks p. 124
- 6.3.4 Reading tasks p. 125
- 6.4 The promise and threats of automated scoring p. 127
- 6.4.1 Examples of current automated scoring programs p. 127
- 6.4.2 Concerns about automated scoring p. 129
- 6.5 Test feedback and reporting p. 131
- 6.6 Online and computer-based resources for statistics p. 135
- 6.6.1 Microsoft“ Excel p. 135
- 6.6.2 Online resources p. 137
- 6.7 Conclusion p. 139
- Further reading p. 139
- Exercises p. 140
- Afterword: the rubber ruler revisited p. 144
- References p. 146
- Index p. 153