

ROUTLEDGE INTRODUCTIONS  
TO APPLIED LINGUISTICS

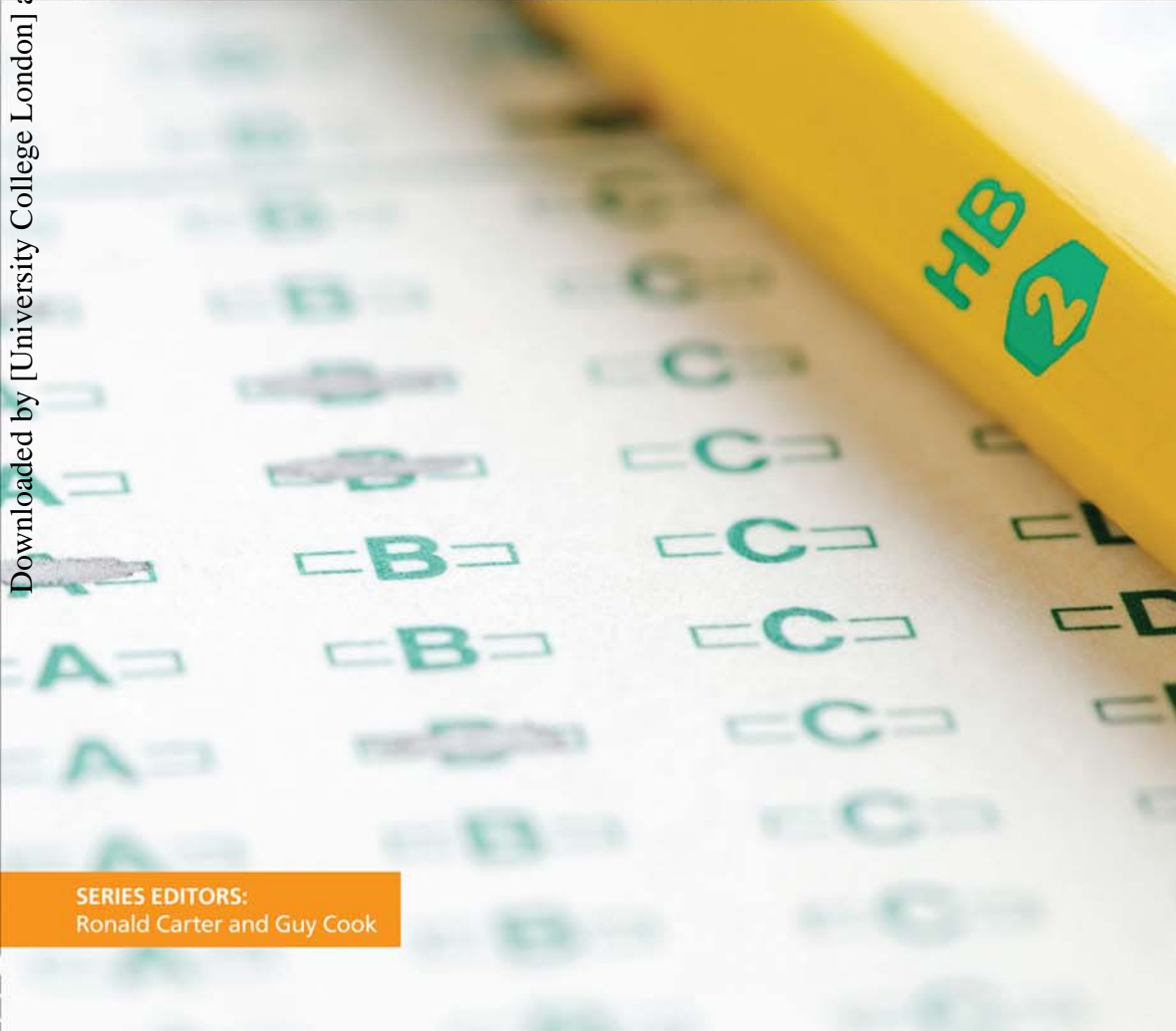


# Exploring Language Assessment and Testing

Language in Action

Anthony Green

Downloaded by [University College London] at 01:47 02 April 2017



SERIES EDITORS:  
Ronald Carter and Guy Cook

# Exploring Language Assessment and Testing

*Routledge Introductions to Applied Linguistics* is a series of introductory level textbooks covering the core topics in Applied Linguistics, primarily designed for those beginning postgraduate studies, or taking an introductory MA course as well as advanced undergraduates. Titles in the series are also ideal for language professionals returning to academic study.

The books take an innovative ‘practice to theory’ approach, with a ‘back-to-front’ structure. This leads the reader from real-world problems and issues, through a discussion of intervention and how to engage with these concerns, before finally relating these practical issues to theoretical foundations. Additional features include tasks with commentaries, a glossary of key terms, and an annotated further reading section.

*Exploring Language Assessment and Testing* is a straightforward introduction to the field that provides an inclusive and impartial survey of both classroom-based assessment by teachers and larger-scale testing, using concrete examples to guide students into the relevant literature.

Ranging from theory to classroom-based scenarios, the author provides practical guidance on designing, developing and using assessments, with flexible step-by-step processes for improving the quality of tests and assessment systems to make them fairer and more accurate.

This book is an indispensable introduction to the areas of language assessment and testing, and will be of interest to language teachers as well as postgraduate and advanced undergraduate students studying Language Education, Applied Linguistics and Language Assessment.

**Anthony Green** is Reader in Language Assessment at the University of Bedfordshire, UK. His other publications include *IELTS Washback in Context* (2007) and *Language Functions Revisited* (2012). He has taught courses in Language Assessment around the world and has extensive practical experience of testing and assessment – including working as a test researcher, developer, item writer and as a language teacher.

# Routledge Introductions to Applied Linguistics

*Series editors:*

**Ronald Carter**, *Professor of Modern English Language,  
University of Nottingham, UK*

**Guy Cook**, *Chair of Language in Education,  
King's College London, UK*

*Routledge Introductions to Applied Linguistics* is a series of introductory level textbooks covering the core topics in Applied Linguistics, primarily designed for those entering postgraduate studies and language professionals returning to academic study. The books take an innovative 'practice to theory' approach, with a 'back-to-front' structure. This leads the reader from real-world problems and issues, through a discussion of intervention and how to engage with these concerns, before finally relating these practical issues to theoretical foundations. Additional features include tasks with commentaries, a glossary of key terms and an annotated further reading section.

## **Exploring English Language Teaching**

Language in Action  
*Graham Hall*

## **Exploring Professional Communication**

Language in Action  
*Stephanie Schmurr*

## **Exploring Classroom Discourse**

Language in Action  
*Steve Walsh*

## **Exploring Language Pedagogy through Second Language Acquisition Research**

*Rod Ellis and Natsuko Shintani*

## **Exploring Corpus Linguistics**

Language in Action  
*Winnie Cheng*

## **Exploring Vocabulary**

Language in Action  
*Dee Gardner*

## **Exploring World Englishes**

Language in a Global Context  
*Philip Seargeant*

## **Exploring Intercultural Communication**

Language in Action  
*Zhu Hua*

## **Exploring Health Communication**

Language in Action  
*Kevin Harvey and  
Nelya Koteyko*

## **Exploring Language Assessment and Testing**

Language in Action  
*Anthony Green*

# Exploring Language Assessment and Testing

Language in Action

Anthony Green

First published 2014  
by Routledge  
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge  
711 Third Avenue, New York, NY 10017

*Routledge is an imprint of the Taylor & Francis Group, an informa business*

© 2014 Anthony Green

The right of Anthony Green to be identified as author of this work has been asserted by him in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

*Trademark notice:* Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

*British Library Cataloguing in Publication Data*

A catalogue record for this book is available from the British Library

*Library of Congress Cataloging in Publication Data*

Green, Anthony, 1966–

Exploring language assessment and testing : language in action / Anthony Green.  
pages cm. – (Routledge introductions to applied linguistics)

Includes bibliographical references and index.

1. Language and languages–Ability testing. 2. Language and languages–Study and teaching. I. Title.

P53.G675 2014

418.0076–dc23

2013010140

ISBN: 978-0-415-59723-4 (hbk)

ISBN: 978-0-415-59724-1 (pbk)

ISBN: 978-1-315-88962-7 (ebk)

Typeset in Sabon  
by Saxon Graphics Ltd, Derby

**For Richard, Maria and Alexander**

This page intentionally left blank

# Contents

<i>List of illustrations</i>	x
<i>Permissions</i>	xii
<i>Acknowledgements</i>	xiv
<i>Series editors' introduction</i>	xv

<b>PART I</b>	<b>1</b>
<b>1 Putting language assessment in its place</b>	<b>3</b>
Language as evidence about people	3
Assessment and testing	6
<b>2 Purposes for assessment</b>	<b>12</b>
Purposes for assessment	12
Educational assessment in teaching and learning	13
<b>3 The practice of language assessment</b>	<b>18</b>
Building quality systems for language assessment	18
Effective assessment comes at a PRICE	21
The assessment cycle	25
<b>4 The qualities of effective assessment systems</b>	<b>58</b>
Four qualities of useful assessments	58
Practicality	59
Reliability	63
Validity	75
Interpretative arguments: assessing the evidence	83
Assessments as maps	85
Beneficial consequences	85
Summary: the right tool for the job	93



<b>PART II</b>	<b>95</b>
<b>5 Assessing receptive skills</b>	<b>97</b>
Defining reading and listening abilities	98
Specifications	106
Assessing grammar and vocabulary	110
Preparing items for assessing reading and listening	111
Trying out material	115
Scoring performance	117
Score reporting and feedback	118
Standard setting	119
<b>6 Assessing productive and interactive skills</b>	<b>127</b>
Defining productive and interactive language abilities	127
Purposes for assessment	135
Assessment design	135
Speaking tasks	138
Task review	140
Trying out material	142
Scoring performance	142
Score reporting and feedback	158
Standard setting	160
<b>PART III</b>	<b>169</b>
<b>7 Multiplication and division: trends in language assessment</b>	<b>171</b>
Introduction	171
What to test? Constructs	173
The psycholinguistic–socio-linguistic phase	184
Integrative testing	188
The mastery learning movement	192
Two approaches to score interpretation	194
Pragmatic tests and the unitary competence hypothesis	196
The ‘promised land’ of communicative testing	198
Rapprochement: communicative language ability	202
Developments in formative assessment	206
Assessment wars: teacher assessment versus external tests	211
Ethical and critical language assessment	214
Higher, wider, deeper: some emerging trends in language assessment	216
Last words	220
<b>8 Commentary on selected tasks</b>	<b>221</b>

<i>Glossary</i>	227
<i>Further Reading</i>	244
<i>References</i>	249
<i>Index</i>	260

# Illustrations

## Figures

3.1	The assessment cycle: roles, responsibilities and documentation	27
3.2	The assessment production cycle	42
4.1	Assessment qualities	58
4.2	Distribution of scores awarded to a Latin composition by 28 expert raters	71
4.3	Inferential bridges in the interpretation of assessment results	83
4.4	Washback causes and effects	87
4.5	The ancient water clock: a simple feedback system	90
5.1	Types of reading based on Urquhart and Weir (1998)	99
5.2	An outline model of receptive language processing based on Weir (2005a) and Field (2008)	101
5.3	Exploring the Arctic (sample reading test material)	123
5.4	Versant Pro – Writing (sample test material)	126
6.1	A simplified model of language production	128
6.2	Interaction in the oral proficiency interview (OPI)	138
6.3	Interaction in the semi-direct or computer-based oral proficiency interview (SOPI/COPI)	139
6.4	Interaction in the paired/group speaking assessment	139
6.5	Example of a simple four-level EBB scale for essay writing based on Upshur and Turner	154
6.6	Writing Task 1 from Versant Pro – Writing	161
6.7	Writing Task 2 from the Pearson Test of English General: Level 3	163
6.8	Speaking Task 1 from <i>Cambridge English: Movers</i>	164
6.9	Speaking Task 2 from the Versant Aviation English Test	165
6.10	Speaking Task 3 from <i>Cambridge English: Business</i>	167
7.1	Auditory comprehension	183
7.2	Sample items from the TOEFL test (from 1976), ‘Listening Comprehension’	184
7.3	Phonetic contrast – the test taker hears: ‘Let me see the sheep’	186

7.4	Sample items from the ELTS <i>Specimen Materials</i> booklet	201
7.5	The Bachman model of language knowledge	203

## Tables

3.1	An example table of specification	34
3.2	Stages of the assessment cycle	56
4.1	Practical considerations in the assessment cycle	62
4.2	Characteristics of assessees	69
4.3	Inferential bridges: steps in building a validity argument for a test of language in the workplace	84
4.4	Characteristics of effective feedback	92
6.1	Features of more and less proficient speech	131
6.2	A basic checklist for assessing short interview tasks	144
6.3	A simple rating scale	145
6.4	A holistic scale: the TOEFL Independent Writing Scale	147
6.5	Qualitative aspects of spoken language use	148
6.6	Comparison between holistic and analytic scoring approaches	150
6.7	Differences between ability and behaviour-focused scales	151
7.1	Phases (or tendencies) in language assessment	174
7.2	The format of the 1913 version of the Certificate of Proficiency in English	175
7.3	Grid for measuring language abilities from Carroll (1961)	179
7.4	The 1964 TOEFL	183
7.5	The design of the ELTS test	200
7.6	The TOEFL iBT and 2012 IELTS tests compared	205
7.7	Interventionist and interactionist assessment for learning	208

# Permissions

The publishers and I would like to thank the following copyright holders for permission to reprint material:

The Educational Testing Service for the sample listening item from the TOEIC test on p. 126, sample items from the TOEFL test (1976) on p. 184 and for the TOEFL Independent Writing Scale on p. 147. All other information contained within this publication is provided by Routledge. No endorsement of any kind by Educational Testing Service should be inferred.

Oxford University Press for extracts from *Oxford Applied Linguistics: Language Assessment in Practice* by Lyle Bachman and Adrian Palmer (2010).

Taylor and Francis for extracts from 'Evolution of a test item' by Mary Spaan from *Language Assessment Quarterly*, vol 4, issue 3 (2007), [www.informaworld.com](http://www.informaworld.com).

Pearson Education Ltd for extracts from *Language Testing: The Construction and Use of Foreign Language Tests* by Robert Lado (1961) and sample items from the *Pearson Test of English General*.

Pearson Education Inc. for extracts from the Versant Pro Writing Test (2012) and Versant Aviation English Test (2012).

Cambridge English Language Assessment for extracts from *Examining Speaking: Research and Practice in Assessing Second Language Speaking* edited by Lynda Taylor (2011); *IELTS Washback in Context: Preparation for Academic Writing in Higher Education* by Anthony Green (2007); *Preliminary English Test Information for Candidates* (2006); *Cambridge Young Learners Tests Sample Papers* (2010); *Cambridge English Business English Certificates Handbook for Teachers* (2012); *Continuity and Innovation: Proficiency in English Examination 1913–2002* edited by Cyril Weir and Michael Milanovic (2003).

The British Council for extracts from *English Language Testing Service Specimen Materials Booklet* (1987).

The Language Policy Division of the Council of Europe for the scale for 'Qualitative aspects of spoken language use' from the *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (2001).

It has not always been possible to identify the sources of all the material used and in such cases the publishers would welcome information from the copyright holders.

# Acknowledgements

This book would never have been possible without the inspiration, help and support of many people. First, I am very grateful to the series editors, Ronald Carter and Guy Cook, for giving me the opportunity and for their encouragement and advice. Thanks also to Emma Hudson, Sophie Jaques and Andrea Service at Routledge for their enthusiastic help in guiding the book along to publication.

My thanks also go to Nick Saville, John Field and Liz Hamp-Lyons, and to other anonymous reviewers who read all or part of the manuscript and offered insightful suggestions; to Cyril Weir, Roger Hawkey and Barry O'Sullivan who have done much to shape my ideas on language assessment; to Lynda Taylor, Fumiyo Nakatsuhara, Sathena Chan, Stephen Bax and Rebecca Selvaratnam at the University of Bedfordshire for their support; to Jin Yan, Masashi Negishi, Alan Davies, Dianne Wall, Caroline Browne, colleagues involved in the ProSET project, especially Natalia Burenina and Olga Safonkina; and to the many others who have helped with enquiries or suggestions. My thanks also must go to the wider community of language educators and assessment professionals around the world whom this book has drawn from.

I am, of course, indebted to my doctoral students, especially Carolyn Westbrook and Daniel Waller, and the many other students and trainees over the years, too many to mention by name, who have raised those difficult or unanswerable questions that have helped me to refine my thinking.

I very much appreciate the people and institutions that have contributed or helped me to source materials and to track down copyright holders. Special thanks are due to Martin Eayrs, to Alistair van Moere of Pearson Knowledge Technologies, Ed Hackett and Alison Sharpe at Oxford University Press, Fiona Barker and John Savage at Cambridge English, Eileen Tyson, Xiaoming Xi and Anthony Ostrander at the Educational Testing Service.

Of course, involvement in a project like this leads to the neglect of other commitments and adds to the burdens of others. The book would never have been possible without the patience and unfailing support of my family. Thanks are especially due to Sachiyo not only for taking on all the extra chores, but for encouraging me throughout.

# Series editors' introduction

## The Introductions to Applied Linguistics series

This series provides clear, authoritative, up-to-date overviews of the major areas of applied linguistics. The books are designed particularly for students embarking on Masters level or teacher-education courses, as well as students in the closing stages of undergraduate study. The practical focus will make the books particularly useful and relevant to those returning to academic study after a period of professional practice, and also to those about to leave the academic world for the challenges of language-related work. For students who have not previously studied applied linguistics, including those who are unfamiliar with current academic study in English-speaking universities, the books can act as one-step introductions. For those with more academic experience, they can also provide a way of surveying, updating and organising existing knowledge.

The view of applied linguistics in this series follows a famous definition of the field by Christopher Brumfit (1995: 27) as:

The theoretical and empirical investigation of real-world problems in which language is a central issue.

In keeping with this broad problem-oriented view, the series will cover a range of topics of relevance to a variety of language-related professions. While language teaching and learning rightly remain prominent and will be the central preoccupation of many readers, our conception of the discipline is by no means limited to these areas. Our view is that while each reader of the series will have their own needs, specialities and interests, there is also much to be gained from a broader view of the discipline as a whole. We believe there is much in common between all enquiries into language-related problems in the real world, and much to be gained from a comparison of the insights from one area of applied linguistics with another. Our hope therefore is that readers and course designers will not choose only those volumes relating to their own particular interests, but use this series to construct a wider knowledge and understanding of the field, and the many



crossovers and resonances between its various areas. Thus, the topics to be covered are wide in range, embracing an exciting mixture of established and new areas of applied linguistic enquiry.

### The perspective on applied linguistics in this series

In line with this problem-oriented definition of the field, and to address the concerns of readers who are interested in how academic study can inform their own professional practice, each book follows a structure in marked contrast to the usual movement *from* theory *to* practice. In this series, this usual progression is presented back to front. The argument moves *from* Problems, *through* Intervention, and *only* finally to Theory. Thus each topic begins with a survey of everyday professional problems in the area under consideration, ones which the reader is likely to have encountered. From there it proceeds to a discussion of intervention and engagement with these problems. Only in a final section (either of the chapter or the book as a whole) does the author reflect upon the implications of this engagement for a general understanding of language, drawing out the theoretical implications. We believe this to be a truly *applied* linguistics perspective, in line with the definition given above, and one in which engagement with real-world problems is the distinctive feature, and in which professional practice can both inform and draw upon academic understanding.

### Support to the reader

Although it is not the intention that the text should be in any way activity driven, the pedagogic process is supported by measured guidance to the reader in the form of suggested activities and tasks that raise questions, prompt reflection and seek to integrate theory and practice. Each book also contains a helpful glossary of key terms.

The series complements and reflects the *Routledge Handbook of Applied Linguistics*, edited by James Simpson, which conceives and categorises the scope of applied linguistics in a broadly similar way.

Ronald Carter  
Guy Cook

### Reference

- Brumfit, C. J. (1995) 'Teacher professionalism and research', in G. Cook and B. Seidlhofer (eds) *Principle and Practice in Applied Linguistics*. Oxford, UK: Oxford University Press, pp. 27–42.

**Note**

There is a section of commentaries on a number of the tasks at the back of the book. The **TC** symbol in the margin indicates that there is a commentary on that task.

This page intentionally left blank

# Part I

**Part I** of this book is intended as a practical resource to help teachers, students, trainees and anyone interested in language education to understand more about the practice of language assessment and testing. A better understanding of the qualities of good assessments and of processes and techniques for improving this should help practising teachers and other educators to make more effective use of assessment in their professional lives.

The first two chapters set out the core functions of language assessments. **Chapter 1** outlines what is meant by language assessment and its place in language education. **Chapter 2** considers some of the many reasons that people have for assessing language abilities. It shows how the practice of language assessment needs to be closely linked to the ways in which results will be used and interpreted. In educational settings, this means there must be an interdependent relationship between teaching, learning and assessment.

**Chapter 3** explains a practical and widely applicable approach to building quality assessments for use in the classroom or in testing programmes. **Chapter 4** describes the qualities of **useful** assessments and ways of evaluating them. These chapters introduce the reader to ways of working that have been shown to improve the quality of assessment materials and procedures.

**Part II** introduces the reader to relevant theoretical insights and shows how the processes described in **Part I** are used in developing tests and assessments. **Chapters 5** and **6** offer straightforward guidance on designing and developing more effective assessments and preparing appropriate materials for the purpose. Supplementary resources on the companion website ([www.routledge.com/cw/rial](http://www.routledge.com/cw/rial)) illustrate the wide range of options open to writers of assessments for learners of all ages and at all stages of language learning.

**Part III** introduces the reader to the study of language assessment, tracing its emergence as a distinct field of enquiry. It outlines developing trends and points to areas of controversy and debate. This aspect of the book not only helps to contextualise and deepen understanding of the practices described, but opens up avenues for the reader to explore through further study.

The focus throughout is on assessing people's ability to use foreign, second or additional languages and reasons for studying this. The principles and techniques described apply to the assessment of any language, but the examples are mainly taken from the assessment of English. This is partly because my own experience is mainly in teaching and assessing English, but also for the very good reason that this book is itself written in English and so the examples should be accessible to all readers.

Technical terms that are covered in the Glossary are identified where they first appear by bold type. Additional resources and practical exercises are provided on the companion website ([www.routledge.com/cw/rial](http://www.routledge.com/cw/rial)). For each section, further sources of reading are recommended for the reader who wishes to explore the issues in more depth.

# 1 Putting language assessment in its place

## Personal reflection

What do you understand by the term *language assessment*? How is it different from *language teaching* or *language learning*?

What do you think are the main differences (if any) between *assessments*, *tests* and *examinations*?

What do you understand by the terms *language skills*, *language knowledge* and *language abilities*?

Think about an experience you have had of a language test.

Did you find the experience enjoyable? Why (or why not)?

Do you think that the test helped you to learn the language? How?

## Language as evidence about people

It is a natural reaction to hearing or reading any fragment of language to use it as evidence about the person who produced it. When we speak on the telephone to people whom we have not met, we may use the sound of their voice, its accent and pitch to give us clues about whether they are native speakers of our language, which region they come from, their gender, age, social class and profession, and their mood and attitudes.

It seems that we can't help forming an opinion about people according to the kind of language they produce. In Britain, surveys of customers of telephone call centres investigating attitudes towards accents show that British people tend to associate American accents with high levels of success, Scottish accents with trustworthiness, and the accents of people who come from large cities such as London, Liverpool and Birmingham with low levels of both (*The Yorkshire Post*, 2006). These attitudes may be unreasonable, but we are more likely to listen to and believe messages delivered by voices that we find pleasing and trustworthy. Businesses know this and carefully choose actors with suitable accents to deliver their advertising messages.

If we receive a letter or email from someone, we base similar judgements on their choice of words and written style. We may even begin to form ideas about their appearance and personality. As you read this, you are probably already building up a picture in your mind of the writer of this book. When you do this kind of thing, you are using the language that a person produces to make **inferences** or draw conclusions about them.

Linguistic evidence can sometimes be used quite formally to give insights into a person's identity or personality. Experts in forensic linguistics often try to establish where a suspect comes from by studying his or her accent or speech patterns. Sometimes employers look at the size and shape of job applicants' handwriting because this is supposed to give insights into their personalities. However, these kinds of inferences about people's regional origins or characteristics are not usually what is implied by the term *language assessment*. In this book the focus is on assessments where knowledge about language, the ability to use language or skill in applying these are measured or evaluated.

In everyday life, we not only make judgements about people on the basis of how they use language; we also make judgements about the qualities of the language that they use. We notice when they make grammatical errors or choose an inappropriate word; when they appear impolite or overly formal; when they are very hesitant; when they are being unclear or imprecise or, on the other hand, when they seem to express their ideas particularly well. We also judge our own use of language. As we speak or write, we monitor and correct ourselves. We put right our slips of the tongue and fix our spelling mistakes (with or without the help of our word processing software); we rephrase the things we say if people misunderstand us; we reorganise our written texts to try to make our ideas clearer to our readers.

The nature of the **knowledge**, **skills** and **abilities** involved in learning and using a language is a matter of debate. An introduction to some of the contentious issues that have excited language assessment specialists is given in **Part III**. Briefly, **knowledge** about language may include recognising a word written in a foreign language and knowing a translation equivalent (*bird* in English can be translated as طائر in Arabic, 禽流 in Chinese or *Vogel* in German), or knowing a grammatical rule: the past participle in English regular verbs is formed by adding *-ed* to the stem: look – looked), or knowing pragmatic conventions. For example, in most circumstances *How are you?* spoken by an acquaintance is likely to be intended as a conventional greeting. The anticipated response might be, *Good. You?* rather than details of the other's feelings or of any medical conditions.

A distinction is often made in language education following Hymes (1972) between knowledge of the rules governing language as a system and the **ability** to actually use language in unrehearsed interaction. Learners may know a good deal about a language, but be unable to

access the words or phrases they know in the course of a conversation or be unable to pronounce them accurately. Conversely, many people are able to communicate effectively in a wide range of situations with little or no explicit knowledge of the systems of rules that govern the language they are using. Language **skills** involve drawing on language knowledge and language abilities in order to read, listen, write, speak, to interact with others, or to mediate between them.

The evidence we have of a person using a language may be very limited – a few telephone conversations and a handful of emails, perhaps – but based on what we observe in these few instances, we often make inferences about their more general knowledge of a language, their ability to use the language and their skill in carrying out language-related **tasks**. We may hear them speak for just a few minutes, but reach conclusions such as: ‘Adel knows more Albanian than I do’ or ‘Bonita can get by in Bengali’. We see a few examples of their writing and we make statements such as: ‘Carole can write very polished business letters in Chinese’ or ‘Dai finds it really difficult to form questions in Danish’.

The major focus of this book will be on the use of language assessment in educational contexts. This setting for language assessment is unusual in some important ways. Firstly, language teachers, unlike most of us, are actually expected to voice their judgements about other people’s language use and are generally rewarded (rather than resented) for pointing out and correcting errors. If you have worked as a language teacher, you will probably have experience during the natural course of any lesson of deciding whether one of your students has produced accurate and appropriate language, and perhaps putting right mistakes they have made or asking them to try again. You have probably asked them to read or listen to a text in the foreign language and then asked them questions about it to check their understanding. As I define language assessment in this book, examples can range from informal classroom activities to well-known national and international language tests.

So, here is a simple definition:

*Language assessment involves obtaining evidence to inform inferences about a person’s language-related knowledge, skills or abilities.*

The *evidence* comes from the performance of tasks that involve the use of language. The *inferences* are what we interpret the performance to mean based on our beliefs about the nature of language and its role in the life of the person being assessed.

The inferences we make about **assesseees** (the term I will use in this book for the people being assessed) are generally used to inform **decisions** – decisions that can range in magnitude from whether it would be better for them to use another comma in a sentence or



practice saying a certain phrase again, to whether they should be given a job or permitted to make their home in a certain country.

In language education, teacher training programmes tend to give assessment rather little attention. Assessment is often the topic of just a few sessions: apparently an issue of little direct concern to the teacher, perhaps better left to expert specialists. On the other hand, it sometimes seems to take on overriding importance in the lives of teachers and learners. Many school children, if asked why they are learning a foreign language, would answer that they are doing it mainly to pass a public test, often known as an **examination**. Most language teachers spend a good deal of time assessing their students and, especially when the day of an important national examination is getting close, many dedicate most of their classes to preparing their students to take these tests.

In this book I argue that language assessment is inseparable from the teaching and learning of languages. This challenges the more traditional view in teacher education that assessment is a distinct activity: one that is marginal to the main business of the language teacher. In my view, teachers need to develop what is sometimes called **assessment literacy**. They not only need to understand something about how language assessments are made, scored and interpreted by others, but also to be able to make, score and interpret the results of useful assessments themselves. Experience suggests that low levels of assessment literacy can lead to bad educational decisions and to teaching and learning that are less effective than they could be.

### Assessment and testing

The traditional view that separates assessment from teaching may result from the identification of the word assessment with the narrower activity of **testing**. A test is an event that is especially set up to elicit a performance (usually within a predetermined time frame) for the purpose of making judgements about a person's knowledge, skills or abilities. In the course of a language test the person being judged (the assessee – a 'test taker', 'testee' or 'candidate') will respond to a **prompt** (e.g., an essay title; questions asked by an interviewer; a set of choices on a test paper). The test taker's response is judged or scored according to a mark scheme. This is a predetermined procedure such as identifying and calculating the proportion of correct answers: 14 out of 20; counting the number of errors in a piece of writing; assigning points or marks to an essay to match descriptions presented on a scale, etc. The outcome is a **score** or **grade**, which is then formally recorded. The formalities and rituals of formal testing and the penalties associated with failure can make them an intimidating experience for the test taker.

For some commentators, assessment is distinguished from testing because it covers a much broader *cycle* of activities. In addition to the

test event (which involves eliciting a performance as evidence of language abilities), these activities include:

- deciding on the content of the test;
- scoring the performance;
- deciding on the meaning of the scores obtained;
- decisions that the scores are used to justify (such as choosing which students to admit onto a course or deciding whether to award certificates).

In this view, the test is just one step in a sequence of events which together make up a **cycle of assessment**. This assessment cycle will be discussed in more detail in [Chapter 3](#).

From another point of view, the word ‘testing’ can refer to this entire cycle of assessment events, but tests are understood to make up a relatively small set of controlled procedures among a much broader *range* of options. Assessment is a more general term than testing and takes in many different methods of obtaining and evaluating language data, including less formal procedures with fewer controls and restrictions than tests.

While it can be helpful to remind ourselves that assessment involves a whole cycle of inter-related activities, it is the second of these distinctions – the range of options for obtaining information – that is more often made in discussions of language assessment and so is the one used in this book.

In contrast to tests, other forms of assessment may involve such activities as:

- informal questioning in class by teachers;
- semi-formal **exercises** and **quizzes** carried out in class;
- learners reflecting on their own use of language – **self-assessment**;
- learners judging each other’s performance – **peer assessment**;
- the collection of samples of language that have not been elicited according to any fixed or formal plan. Examples of this include **observations** and some forms of **portfolio assessment**.

Observations involve teachers watching and recording student performance in classroom activities. Portfolios are collections of student work that may showcase their best achievements or represent progress over a period of time.

### Task 1.1

Have you ever experienced self-assessment or peer assessment?

What do you think would be the main advantages and disadvantages of these techniques?

Other approaches to assessment that do, like tests, involve specific prompts (e.g., classroom exercises of the kind found in most textbooks) may not be carried out under controlled conditions (such as set time constraints or curbs on discussing answers with fellow students). Scores may not be formally recorded as evidence of the students' abilities. In fact, reporting does not need to involve the award of grades or scores at all, but may, for example, involve highlighting errors or giving descriptive commentaries on performance.

It is important to understand that tests are not the only means of judging learners' language knowledge, skills and abilities; and that not all language assessments are **formal** procedures that lead to scores or grades. On the contrary, most language assessments are **informal**, unobtrusive, involve no special arrangements and do not cause particular anxiety or fear in the people who are being assessed.

### **Assessment, teaching and learning**

There are a number of different strategies we can use when we learn a new skill from another person. Perhaps the simplest of these is imitation: one person, the learner, watches what another person does and then attempts it for him or herself. When animals learn skills, this generally appears to be how they do it. We know that some chimpanzees, for example, can learn how to carry out quite sophisticated tasks, such as using a stick as a tool to catch insects, by copying the actions of other members of their group. They carefully observe what the skilled chimpanzee does, recognising that one set of actions – such as shaping a stick – is related to another – poking the stick into a tree trunk – and that these actions together bring a reward – a meal of insects. They then try to recreate the same series of actions themselves to get the same reward. Chimpanzees are very good at imitation and they often perform better than humans in carrying out intricate imitation tasks.

Assessment of this kind of learning can be straightforward. Combining imitation with a process of trial and error, learners either succeed in accomplishing the task, or repeatedly fail and eventually give up.

Although it is possible to build knowledge by observing and copying, imitation alone is a limited way of passing on skills, and the more complex the skill, the less effective imitation seems to be. A child can watch an adult driving a car and try to copy the movements she sees, but that will probably not make her a safe and effective driver when she has the chance to take the wheel. When learning a language, I can try to imitate the sounds that speakers of that language make, to memorise words and grammatical patterns; but that is not likely to be enough to enable me to communicate effectively.

Learning complex skills like these is much more effective if we understand something about *why* people perform certain actions and how those actions help them to accomplish the tasks they want to carry out. This level of understanding is easier to achieve if we have another person to *teach* us.

In order to teach, people need both to understand that others can't do what they themselves can do, and to be motivated to pass on their own knowledge. When a person has mastered a new skill, they often share their newfound ability not only by showing – inviting others to watch as they perform the skill – or telling – explaining to the learners how to do it – but also by attending to the learners and judging how well they have understood and how close they are to being able to carry out the task independently.

Human learners do not usually simply imitate their teacher, but actively try to make sense of what they are learning: not only to notice that carrying out a sequence of actions leads to a particular outcome, but to understand *why* it does so. The teacher can support the learner by observing the learner's efforts and providing **feedback**: pointing out what they are doing well and what they are doing poorly in order to help them to improve. As humans, our willingness to teach each other in this way is one reason why we are so much more efficient than animals at preserving and developing our collective knowledge and technologies.

Understanding the gap between what learners can do now and what they need to be able to do or understand in order to fulfil a task successfully requires a sophisticated awareness both of the task and the learner. In the case of language learning, learners need to build an understanding of how people can use language to accomplish tasks such as establishing a rapport with another person, buying clothes or getting their hair cut. Assessment that takes account of this requires more than simple judgements of success or failure.

As well as having a good mastery of the skill herself, and the ability to model the skill in ways that make it more accessible to the learner, the good teacher breaks down complex skills into the different elements that contribute to success. The teacher recognises which of these the learners are able to do independently, which they can do with some help or prompting, and which remain well beyond their abilities. The teacher may allow learners to carry out parts of a task that they are already able to accomplish, but offer help with the more challenging stages. Think of how a mother might help a child to assemble a model or cook a meal, letting the child carry out the task, but asking questions, pointing out mistakes, perhaps actually carrying out some of the most complex steps in the process herself before asking the child to try to replicate what she has done, giving feedback at the same time on what the child is doing well and how he might do better.

Another way in which people learn new skills is through working together to solve problems. By acting as a team, people can do things that no individual member could do when acting alone. Think of teams of engineers or medical researchers developing innovative machinery or new treatments. The members of the group create new forms of knowledge by participating together in activities and the relationships between them can be very important to the kinds of learning that will occur.

This kind of collective social learning can be challenging to assess, especially when using traditional tests, grades and scores, because it is difficult to establish what contribution each group member has made to the outcome. In order to achieve better outcomes, members of the group may need to come to understand for themselves what aspects of the activity they might be able to carry out in different ways to achieve better results.

Because assessment is used for a wide variety of purposes, it is quite easy to conceive of assessment taking place without teaching – one reason why many learners don't like public tests is that the test providers do not often give them any advice on which questions they answered correctly or incorrectly. The test teaches them nothing. On the other hand, it is almost impossible to imagine any form of teaching that does not involve assessment. Even in the most traditional classrooms, based on a transmission model of teaching (which assumes that learning is mainly or wholly dependent on the teacher telling a group of students new information), the content of the lectures is intended to fill a gap between what the students are believed to know already and what they need to know in order to succeed.

To be really effective as teachers, we need to find effective ways of assessing learners. We then need to use what we learn from assessment to decide on actions to take that will help learners to improve their knowledge of the language or their skill in using it. A major challenge for language educators is that the process of learning a second language is still rather poorly understood. **Second language acquisition** (SLA) researchers, who investigate this process, disagree about the fundamental mechanisms of language learning and a number of competing theories can be found (see Cook, 2008; Lightbown and Spada, 2012; Mitchell *et al.*, 2012). In assessing learners, we need to be clear about how we believe languages are learned; what kinds of evidence we might be able to collect to show us what learners know about a language; what they are able to use the language to do; what they have difficulty with; and how we can help them to expand their abilities.

Another troublesome issue in assessing language learning is the extent of disagreement about the purposes for learning languages. For some, languages should be taught in schools as a necessary basic skill. Just as we need to learn arithmetic in order to operate in the modern

world – shopping, using banking services, checking our taxes – so we need foreign languages to do business and to interact with people from other countries. There are said to be economic benefits for countries with a linguistically adaptable labour force.

For others, the primary benefit of learning languages has more to do with what Cook (2008) described as ‘brain training’. Languages help us to develop our intelligence or adopt more flexible ways of thinking. Traditionally, a major purpose for learning languages was personal development through the understanding of classic works of literature and so becoming more ‘cultured’. What we choose to assess will, of course, reflect what we think are the key purposes for language learning. These issues will be explored in more detail in [Part III](#).

People generally tend to believe that language assessments give them pertinent information that can help them to make good decisions. If they are right, assessments may be useful tools that improve teaching and learning and serve society more generally. On the other hand, a lack of assessment information, poor information from badly conceived assessments, or even poorly understood information from well-conceived assessments may lead to regrettable decisions. Teaching and learning will suffer, able people will be denied opportunities and society will experience the negative effects. It is very important that we carry out language assessment as well as we can and use the results in a well-informed manner.