

CAMBRIDGE

# Testing for Language Teachers

Third Edition

Arthur Hughes and Jake Hughes

**CAMBRIDGE**  
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,  
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of  
education, learning and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781108714822](http://www.cambridge.org/9781108714822)

© Cambridge University Press 2020

This publication is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without the written  
permission of Cambridge University Press.

First published 2020

20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Printed in Great Britain by Ashford Colour Press Ltd.

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

ISBN 978-1-108-71482-2 Paperback

ISBN 978-1-108-71485-3 Apple iBook

ISBN 978-1-108-71483-9 ebooks.com eBook

ISBN 978-1-108-71487-7 Google eBook

ISBN 978-1-108-71486-0 Kindle eBook

Cambridge University Press has no responsibility for the persistence or  
accuracy of URLs for external or third-party internet websites referred to in  
this publication, and does not guarantee that any content on such websites is,  
or will remain, accurate or appropriate.

11	Testing reading	140
12	Testing listening	163
13	Testing grammar and vocabulary	176
14	Testing overall ability	192
15	Tests for young learners	206
16	Beyond testing: other means of assessment	227
17	New technology and language testing	233
18	Test administration	238
19	The statistical analysis of test data	242
	Appendix 1 Item banking	256
	Appendix 2 Checklist for teachers choosing tests for their students	258
	Appendix 3 The secrets of happiness	260
	Bibliography	261
	Author index	278
	Subject index	282

## 4. Informal trialling of items on expert speakers

Items which have been through the process of moderation should be presented in the form of a test (or tests) to a number of expert speakers – twenty or more, if possible. There is no need to do this formally; the ‘test’ can be taken in the participants’ own time. The expert speakers should be similar to the people for whom the test is being developed, in terms of age, education and general background. There is no need for them to be specialists in language or testing. Indeed, it is preferable that they should not be, since ‘experts’ are unlikely to behave in the same way as naïve test-takers.

Items that prove difficult for the expert speakers almost certainly need revision or replacement. So do items where unexpected or inappropriate responses are provided. Of course, people taking a test on their own language will have lapses of attention. Where these can be recognised, the responses should not count against the item.

## 5. Trialling of the test on a group of non-expert speakers similar to those for whom the test is intended

Those items that have survived moderation and informal trialling on expert speakers should be put together into a test, which is then administered under test conditions to a group similar to that for which the test is intended<sup>4</sup>. Problems in administration and scoring are noted.

It has to be accepted that, for a number of reasons, trialling of this kind is often not feasible. In some situations a group for trialling may simply not be available. In other situations, although a suitable group exists, it may be thought that the security of the test might be put at risk. It is often the case, therefore, that faults in a test are discovered only after it has been administered to the target group. Unless it is intended that no part of the test should be used again, it is worthwhile noting problems that become apparent during administration and scoring, and afterwards carrying out statistical analysis of the kind referred to below and treated more fully in Chapter 19.

<sup>4</sup> If there are too many items for one group to take in a single sitting, more than one form of the test can be constructed, with each form containing a subset of items common to both (known as anchor items). Using performance on the common anchor items as a basis for comparison, it is possible to put the other items on the same difficulty scale. If this is not done, differences in ability between the groups will mean that the difficulty levels of items taken by one group will not be directly comparable with the difficulty levels of items taken by another group. See Chapter 19 for statistical treatment of results when anchor items are used.

elements that apply to their own situation. There will be some points where perhaps more detail is called for; others where additional elements are needed. There is certainly no reason to feel limited to this particular framework or its content, but all in all these specifications should provide a good starting point for many testing purposes. For the same reason, further examples of specifications are given in the following chapters.

A second example, this time much more restricted, concerns the writing component of a test of English for academic purposes with which one of us was associated. The purpose of the test was to discover whether a student's written English was adequate for study through the medium of English at a particular overseas university. An analysis of needs had revealed that the most important uses of written English were for the purpose of taking notes in lectures and the writing of examination answers up to two paragraphs in length. The first of these tasks was integrated into the listening component of the test. This left the examination answers. An analysis of examination questions in the university revealed that students were required to describe, explain, compare and contrast, and argue for and against a position. Because in that university the first-year undergraduate course is very general (all students study arts, science and social science subjects), almost all reasonably academic topics were appropriate. The addressees were university lecturers – both expert speakers and non-expert speakers of English. Using the suggested framework, we can describe the relevant tasks quite succinctly:

### **Operations**

Describe, explain, compare and contrast, argue for and against a position.

### **Types of text**

Examination answers up to two paragraphs in length.

### **Addressees of texts**

Expert speaker and non-expert speaker university lecturers.

### **Topics**

Any capable of academic treatment. Not specialist. Relevant to the test-takers.

### **Dialect and style**

Any standard variety of English (e.g. American, British) or a mixture of these. Formal style.

### **Length of texts**

About one page.

## Communicative competence

Roever and Kasper (2018) argue for interactional competence to be incorporated into speaking assessment. Roever (2011) reviews existing tests of pragmatic competence and makes suggestions for future pragmatics tests. Youn (2015) investigates the effectiveness of role play activities in assessing pragmatic competence.



## READER ACTIVITIES

1. Construct items to test the following:

- Conditional: *If ... had ... ... would have ...*
- Comparison of equality.
- Relative pronoun *whose*.
- Past continuous: *... was -ing, when ...*

Which of the techniques suggested in the chapter suits each structure best? Can you say why?

2. Can you see anything wrong with the following multiple choice items taken from tests written by teachers (use the checklist given as Table 2 in Chapter 7). If so, what? Try to improve them.

- a. I said to my friend ' \_\_\_\_\_ be stupid.'

Isn't      Aren't      Didn't      Don't be

- b. What \_\_\_\_\_ you do, if your car broke down?

must      did      shall

- c. You are too thin. You should eat \_\_\_\_\_

many      more      a few

- d. I'm sorry that the child saw the accident.

– I don't think it matters. He soon \_\_\_\_\_ it.

is forgetting      forgets      will forget      will be forgetting

- e. People \_\_\_\_\_ in their reaction to the same stimulus.

replace      vary      upset      very

3. Produce three vocabulary tests by writing three items for each of the following words. One set of items should be multiple choice *without* context; one set should be multiple choice *with* context; the third set should be gap filling. Give each test to a different (but comparable) group of students. Compare performance on items testing the same word. Can differences of performance be attributed to a difference in technique?

beard	sigh	bench	deaf	genial
tickle	weep	greedy	mellow	callow

(If the words are inappropriate for your students, replace them with others.)

4. Connotation and collocation are notoriously difficult to test but they could well form part of the non-testing assessment of vocabulary (see Chapter 16). How would you assess a student's control of connotation and collocation? Give two examples of each.
5. Look at the paraphrase items from the *Cambridge English B2 First Handbook* on pages 181–182. For each item, identify whether it is testing grammar, vocabulary or both. Compare with a colleague.