# Testing for Language Teachers

Second Edition

Arthur Hughes



CAMBRIDGE UNIVERSITY PRESS Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
www.cambridge.org
Information on this title: www.cambridge.org/9780521484954

© Cambridge University Press 1989, 2003

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 1989 Second edition 2003 9th printing 2009

Printed in the United Kingdom at the University Press, Cambridge

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

ISBN 978-0-521-48495-4 Paperback ISBN 978-0-521-82325-8 Hardback

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. Information regarding prices, travel timetables and other factual information given in this work are correct at the time of first printing but Cambridge University Press does not guarantee the accuracy of such information thereafter.

# **Contents**

Acl	acknowledgements				
Preface					
1	Teaching and testing	I			
2	Testing as problem solving: an overview of the book	8			
3	Kinds of tests and testing	II			
4	Validity	26			
5	Reliability	36			
6	Achieving beneficial backwash	53			
7	Stages of test development	58			
8	Common test techniques	75			
9	Testing writing	83			
10	Testing oral ability	113			
11	Testing reading	136			
12	Testing listening	160			
13	Testing grammar and vocabulary	172			
14	Testing overall ability	186			
15	Tests for young learners	199			
		vii			

# Contents

16 Test administration	215	
Appendix 1 The statistical analysis of test data	218	
Appendix 2 Item banking	234	
Appendix 3 Questions on the New Zealand youth hostels passage	236	
Bibliography	237	
Subject Index		
Author Index	250	

Links to all of the websites mentioned in the book can be found there. The increasing tendency throughout the world for language learning (and testing) to begin at primary school has led me to include a chapter on the testing of young learners. While one might have reservations about such testing (as opposed to other forms of assessment), since testing does take place, it seemed better to offer advice in the area rather than simply ignore the issue.

Perhaps the most striking development in language testing since the publication of the first edition has been the great increase in the number of published articles and books. Many – perhaps most – of the articles have been of a theoretical or technical nature, not directly relevant to the concerns of language teachers. Even where their relevance is clear, in order to keep the text accessible to newcomers to the field, I have usually restricted references to them to the Further reading sections. These sections are intended to act as a guide for those readers who wish to go more deeply into the issues raised in the book, and also to provide an outline of the state of language testing today. They also contain recommendations of a number of recent books which, in an accessible fashion, treat areas of language testing (such as the testing of a particular skill) in greater depth than is possible in the present volume.

I must acknowledge the contributions of others: MA and research students at Reading University, too numerous to mention by name, who have taught me much, usually by asking questions that I found difficult to answer; my friends and colleagues, Paul Fletcher, Michael Garman, Don Porter, Tony Woods, who all read parts of the manuscript of the first edition and made many helpful suggestions; Angela Hasselgren who shared thoughts on the testing of young learners and provided me with copies of the materials used in the Norwegian EVA project; my friends Cyril Weir and Russanne Hozayin, with whom I've collaborated on testing projects in recent years; and finally my wife, who drew the cartoon series on page 209, and whose patience during the writing of this second edition was almost endless.

## Testing for language teachers

Which test seems the more reliable? The differences between the two sets of scores are much smaller for Test B than for Test A. On the evidence that we have here (and in practice we would not wish to make claims about reliability on the basis of such a small number of individuals), Test B appears to be more reliable than Test A.

Look now at Table 1(c), which represents scores of the same students on an interview using a five-point scale.

Table 1(	c):	Scores	on	interview	(Invented	data)

Student	Score obtained	Score which would have been obtained on the following day
Bill	5	3
Mary	4	5
Ann	2	4
Harry	5	2
Cyril	2	4
Pauline	3	5
Don	3	1
Colin	1	2
Irene	4	5
Sue	3	1

In one sense the two sets of interview scores are very similar. The largest difference between a student's actual score and the one which would have been obtained on the following day is 3. But the largest possible difference is only 4! Really the two sets of scores are very different. This becomes apparent once we compare the size of the differences between students with the size of differences between scores for individual students. They are of about the same order of magnitude. The result of this can be seen if we place the students in order according to their interview score, the highest first. The order based on their actual scores is markedly different from the one based on the scores they would have obtained if they had had the interview on the following day. This interview turns out in fact not to be very reliable at all.

# The reliability coefficient

It is possible to quantify the reliability of a test in the form of a reliability coefficient. Reliability coefficients are like validity coefficients

Vocabulary range General academic, not specialist technical.

Dialect and style Standard American or British English dialect. Formal, academic style.

Speed of processing Expeditious: 300 words per minute (not reading all words).

Careful: 100 words per minute.

## Structure, timing, medium and techniques

Test structure Two sections: expeditious reading; careful reading.

Number of items 30 expeditious; 20 careful. Total: 50 items.

Number of passages 3 expeditious; 2 careful.

Timing Expeditious: 15 minutes per passage (each passage collected after 15 minutes).

Careful: 30 minutes (passage only handed out after 45 minutes, when expeditious reading has been completed).

TOTAL: 75 minutes.

Medium Paper and pencil. Each passage in a separate booklet.

Techniques Short answer and gap filling for both sections.

# Examples:

For inferring meaning from context:

For each of the following, find a single word in the text with an equivalent meaning. Note: the word in the text may have an ending such as -ing, -s, etc.

highest point (lines 20–35)

For identifying referents:

What does each of the following refer to in the text? Be very precise.

the former (line 43)

# Criterial levels of performance

Satisfactory performance is represented by 80 per cent accuracy in each of the two sections.

The number of students reaching this level will be the number who have succeeded in terms of the course's objectives.

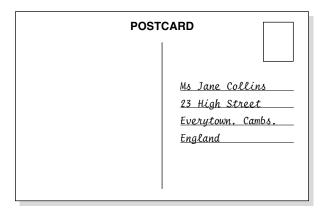
#### TASK 3

You are now working in the American Summer Camps for Children in Florida. You write a postcard to an English-speaking friend.

On your postcard tell your friend:

- where you are
- why you are there
- two things you like about the Summer Camp.

Write your POSTCARD here.



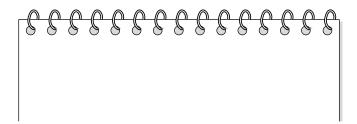
#### TASK 4

You have arranged to go out tonight with Gerry and Carrie, two other Helpers at the Summer Camp in Florida. You have to change your plans suddenly, and cannot meet them. You leave them a note.

In your note:

- apologise and explain why you cannot meet them
- suggest a different day to go out.

Write your NOTE here.



## Testing for language teachers

through higher level items. The same is true for grammar and vocabulary. They are both tested indirectly in every reading test, but the place for grammar and vocabulary items is, I would say, in grammar and vocabulary tests. For that reason I will not discuss them further in this chapter.

To be consistent with our general framework for specifications, we will refer to the skills that readers perform when reading a text as 'operations'. In the boxes that follow are checklists (not meant to be exhaustive) which it is thought the reader of this book may find useful. Note the distinction, based on differences of purpose, between expeditious (quick and efficient) reading and slow and careful reading. There has been a tendency in the past for expeditious reading to be given less prominence in tests than it deserves. The backwash effect of this is that many students have not been trained to read quickly and efficiently. This is a considerable disadvantage when, for example, they study overseas and are expected to read extensively in very limited periods of time. Another example of harmful backwash!

## **Expeditious reading operations**

## Skimming

The candidate can:

- obtain main ideas and discourse topic quickly and efficiently;
- establish quickly the structure of a text;
- decide the relevance of a text (or part of a text) to their needs.

# Search reading

The candidate can quickly find information on a predetermined topic.

# Scanning

The candidate can quickly find:

- specific words or phrases;
- figures, percentages;
- specific items in an index;
- specific names in a bibliography or a set of references.

Note that any serious testing of expeditious reading will require candidates to respond to items without having time to read the full contents of a passage.

What is a college?
Confusion exists concerning the real purposes, aims, and goals of a college. What are these? What should a college be?
Some believe that the chief function 1even a liberal arts college is 2vocational one. I feel that the 3function of a college, while important, 4nonetheless secondary. Others profess that the 5 purpose of a college is to 6 paragons of moral, mental, and spiritual 7 Bernard McFaddens with halos. If they 8 that the college should include students 9 the highest moral, ethical, and religious 10 by precept and example, I 11 willing to accept the thesis.
I 12 in attention to both social amenities 13 regulations, but I prefer to see 14 colleges get down to more basic 15 and ethical considerations instead of standing in loco parentis 16 four years when 17 student is attempting in his youthful 18 awkward ways, to grow up. It 19 been said that it was not 20 duty to prolong adolescences. We are 21 adept at it.
There are those 22 maintain that the chief purpose of 23 college is to develop "responsible citizens." 24 is good if responsible citizenship is 25 by-product of all the factors which 26 to make up a college education 27 life itself. The difficulty arises from 28 confusion about the meaning of responsible 29 I know of one college which 30 mainly to produce, in a kind 31 academic assembly line, outstanding exponents of 32 system of free enterprise.
Likewise, I 33 to praise the kind of education 34 extols one kind of economic system 35 the exclusion of the good portions 36 other kinds of economic systems. It 37 to me therefore, that a college 38 represent a combination of all 39 above aims, and should be something 40 besides – first and foremost an educational 41, the center of which is the 42 exchange between teachers and students.
I 43 read entirely too many statements such 44 this one on admissions application papers: "45 want a college education because I 46 that this will help to support 47 and my family." I suspect that 48 job as a bricklayer would help this 49 to support himself and his family 50 better than a college education.
(Oller and Conrad 1971)