

A language testing handbook

Andrew Harrison

Essential
Language
Teaching
Series

58073
37

General editor:
Roger H Flavell

A Language Testing Handbook

by Andrew Harrison

Essential Language Teaching Series

General Editor: Roger H Flavell



Macmillan Press London

© Text Andrew Harrison 1983

© Illustrations The Macmillan Press Ltd 1983

All rights reserved. No part of this publication
may be reproduced or transmitted, in any
form or by any means, without permission.

First published 1983

Published by

THE MACMILLAN PRESS LIMITED

London and Basingstoke

Associated companies throughout the world

ISBN 0 333 27174 2

Filmset in Lasercomp 11/13pt. Apollo by
August Filmsetting, Warrington, Cheshire

Printed in Hong Kong

List of test types and applications

code letter	test type	placement	diagnostic	achievement	proficiency
A	scripted speech + true/false items	1	3	3	3
B	narrative text + true/false items	1	3	3	3
C	structured writing	1	2	2	2
D	cloze	1	x	2	2
E	dictation	1	2	2	2
F	conversation	1	x	2	2
G	scripted speech + multiple-choice pictures	x	1	3	x
H	scripted speech + completion items	x	1	3	x
J	completion + write	x	1	2	x
K	completion + multiple-choice fillers	x	1	3	x
L	transposition	x	1	2	x
M	unscripted speech + multiple-choice items	2	3	1	2
N	unscripted speech + visuals	2	3	1	1
O	text and argument + multiple-choice items	2	3	1	2
P	letter	2	3	1	2
Q	reorientation	x	2	1	x
R	speak to pictures	2	2	1	3
S	talk on topic	2	x	1	1
T	transfer	3	3	2	1
U	follow instructions	2	2	2	1
V	give advice	x	2	3	1
W	appropriate response	x	3	2	1
X	sequence	x	3	3	1
Y	role play	x	2	2	1
Z	problem solving	x	x	2	1

Note: The numbers indicate how useful each type of test is likely to be for the four purposes, placement, diagnostic, achievement and proficiency, ranging from 1 (most useful) to 3 (useful only in some circumstances) x means not suitable for this purpose

Note

An asterisk in the text means that the word or phrase alongside it is explained in the Glossary beginning on page 140.

Contents

<i>List of test types and applications</i>	v
Introduction	1
1 Types of test	4
1.1 Placement	4
1.2 Diagnostic	6
1.3 Achievement	7
1.4 Proficiency	7
2 Qualities of a good test	10
2.1 Reliability	10
2.2 Validity	11
2.3 Practicality	12
2.4 Comparison and discrimination	13
3 Test design	16
3.1 Outline for a specification	16
4 Placement tests	24
4.1 Review	24
4.2 Commentary	26
4.3 Setting the tests	31
5 Diagnostic tests	49
5.1 Review	49
5.2 Commentary	49
5.3 Setting the tests	52

6	Achievement tests	64
6.1	Review	64
6.2	Commentary	65
6.3	Setting the tests	67
7	Proficiency tests	91
7.1	Review	91
7.2	Commentary	92
7.3	Setting the tests	95
8	Marking	110
8.1	Objective	110
8.2	Subjective	111
8.3	Dictation	114
8.4	Cloze	114
8.5	Speaking	115
9	Results	118
9.1	Distribution and mean	119
9.2	Standard deviation	123
9.3	Reliability	126
9.4	Item analysis	127
9.5	Follow-up	132
	Conclusion	134
	Bibliography	136
	Useful addresses	139
	Glossary	140
	Index	144

Introduction

A common view of testing is that it is quite separate from teaching and learning, both theoretically and in practice. According to this view, a test is a necessary but unpleasant imposition from outside the classroom: it helps to set standards but uses up valuable class time. Underlying this book is the belief that, far from being divorced from each other, testing and teaching are closely interrelated. A test is seen as a natural extension of classroom work, providing teacher and student with useful information that can serve each as a basis for improvement.

The usefulness of the information derived from a test will depend upon the amount of care that is taken in its preparation. There seems to be a widespread belief that good tests are produced only by experts familiar with technical concepts and statistics. However, once it is accepted that testing is an integral part of teaching, it follows that the person best prepared to set the test is the teacher. When testing procedures are related more to teaching and learning than to theories of psychometrics (measurement of the mind) and statistical probabilities, then the experienced teacher is already halfway to being a test setter. What he still has to learn is how the principles and techniques of testing differ from those of teaching, and how to apply them.

This book falls into three parts. The first part (Chapters 1-3) is important for those who have no previous experience of testing, because it explains the concepts and principles on which the rest of the book is based. More experienced testers, however, should also find it useful as an introduction to the second part (Chapters

4-7), which offers a wide range of ideas for different types of test. The intention here is not to give a complete survey of all varieties of language test: this ground is already well covered by other books (see bibliography, page 136). The aim is to provide guidelines for setting several kinds of test which will be practical in use and give helpful information to both teachers and students about their successes and failings. The examples which illustrate the discussion are mainly invented for the purpose and are not meant to be used directly in the classroom. The reason for this is that a good test, like a good suit, should be made for the individual customer ('made to measure' is perhaps too ambiguous a phrase in the present context). The teacher must choose among the tests and techniques available, and adapt them to fit his particular situation. Finally, Chapter 8 takes up matters related to marking, while Chapter 9 gives a brief account of procedures which will help teachers to find out how well their tests have worked and to interpret students' scores.

It will be clear by now that the book is concerned with tests set by teachers for their own students, and not with external examinations such as those set by the Cambridge Syndicate or the Royal Society of Arts. The starting point for the tests in this book is always teaching and learning, with the assumption that the teacher's aim in the long run is to equip his students not with a general knowledge of grammar and vocabulary but with the particular language skills that they will need as, for example, a tourist or postgraduate. This means that the principles of the communicative approach to language teaching are never far from the surface, although the word 'communicative' itself is little used in the text.

A consequence of this approach is that each test is designed as far as possible from the point of view of the student taking it. Once the various skills needed by the student to do the test are properly set out, it becomes clear that test descriptions like 'reading comprehension' and 'oral' are not sufficient to indicate what the

test is assessing. The titles given to the tests described in Chapters 4-7 may seem unconventional, but they are an attempt to show that a test does not assess skills in isolation from each other, or even in pairs, but in complex interactions which the student must cope with in order to complete the task set by the test.

1 Types of test

One of the arguments for an approach to testing through teaching and learning is that the teacher is in the best position to know which tests are appropriate for his class. The appropriateness of a test is largely determined by purpose: why is a test needed at a particular stage in the students' learning and what use will be made of the results? The four types of test described in this book are: placement, diagnostic, achievement and proficiency, though the categories of test and the names given to them vary considerably according to the preferences of different writers. Figure 1 shows the general content, the purpose and the factors which have to be considered for each of the four categories.

1.1 Placement

A *placement* test is designed to sort new students into teaching groups, so that they can start a course at approximately the same level as the other students in the class. It is concerned with the student's present standing, and so relates to general ability rather than specific points of learning. As a rule the results are needed quickly so that teaching may begin. This puts a severe constraint upon the types of test that can be used. At the same time a variety of tests is necessary because a range of different activities is more likely to give an accurate overall picture of a student's level than a single assessment

Figure 1 A framework for language assessments

category	content	purpose	considerations
placement	general reference forward to future learning	grouping	speed of results variety of tests interview
diagnostic	detailed reference back to classwork	motivation remedial work	short term objectives new examples of the material taught
achievement	general reference back to course	certification comparison with others at the same stage	decisions about sampling similar material to that taught in new contexts
proficiency	specific purposes reference forward to particular applications of language acquired	evidence of ability to use language in practical situations	definition of operational needs authenticity context strategies for coping

It is also advisable for one member of staff to see each student individually, if only for two or three minutes, before the final class allocation is made. (This is, of course, vital if the course deals primarily with the spoken language.) At this interview the member of staff can coordinate the results of the various tests that the student has just taken, and add some personal notes to the student's file. This procedure has several advantages. It helps to complete the assessment for each individual student by disclosing factors which are not revealed by the written tests, either positive ones such as a friendly, outgoing character or a higher level of production than a writing test suggests, or negative ones such as a slight stammer or more than average shyness. Also, it establishes each student right from the start of his course as an individual, not just a name and number in the filing system. At a practical level,

the interview can sort out minor problems related to application forms and registration procedures. This provides a genuine purpose for the use of the spoken language, which achieves its own end while giving the 'examiner' material on which to assess the student's oral capability.

Perhaps the greatest advantage of the interview is that there is now the opportunity to assess both oral production (the ability to make English sounds) and fluency (the ability to sound English in a social situation) at one and the same time.

1.2 Diagnostic

A *diagnostic* test (sometimes called a *formative* or *progress* test) checks on students' progress in learning particular elements of the course. It is used for example at the end of a unit in the course-book or after a lesson designed to teach one particular point. These tests can take the form of an extension of the lesson from a practice phase into an assessment phase. Carefully organised, however, they can provide information about progress which may be used systematically for remedial work. The diagnostic test tries to answer the question 'How well have the students learnt this particular material?' Since it relates to particular elements in the course which have just been taught, for example 'type 3 conditional sentences with *if*' or 'asking permission', the assessment will give immediate feedback to the student. If his learning has been successful, the results will give a considerable lift to the student's morale and he is likely to approach the next learning tasks with fresh enthusiasm. If he finds he has not mastered the point at issue, the test should give him clear indications of how he falls short, so that he can do some useful revision. The demands for this kind of test are therefore that it must relate to specific short-term objectives and that it should include further examples of the same kind of material as that used in teaching.

1.3 Achievement

An *achievement* test (also called an *attainment* or *summative* test) looks back over a longer period of learning than the diagnostic test, for example a year's work, or a whole course, or even a variety of different courses. It is intended to show the standard which the students have now reached in relation to other students at the same stage. This standard may be worldwide, as with the Cambridge examinations in EFL; or established for a country, as with school-leaving certificates; or it may relate to an individual school or group of schools which issues certificates to students attending courses. But the important point which is common to all these situations is that the standard remains constant as far as possible from course to course and from year to year and is external to the individual class or textbook. The conditions for setting an achievement test are that it covers a much wider range of material than a diagnostic test and relates to long-term rather than short-term objectives. This brings up problems of sampling, since what has been learnt in a year (for example) cannot all be assessed in one day, yet the test must reflect the content of the whole course. Decisions therefore have to be made about what should be included in the test, and whether assessing one thing can be assumed to include another. For example, if a student can cope with the form and meaning of the past perfect tense, does this imply a similar mastery of the present perfect, since the normal sequence of learning deals with the second of these before the first? On a wider scale, if the student has learnt to write a business letter, can it be assumed that he can cope with personal letters too, on the same basis of the normal sequence of learning?

1.4 Proficiency

The aim of a *proficiency* test is to assess the student's ability to apply in actual situations what he has learnt. It seeks to answer

the question: 'Having learnt this much, what can the student do with it?' This type of test is not usually related to any particular course because it is concerned with the student's current standing in relation to his future needs. In view of this future orientation, a proficiency test is the most suitable vehicle for assessing English for Specific Purposes (ESP). However, the purposes for which it serves as an appropriate testing instrument extend far beyond those which are normally cited as 'specific purpose' (such as language for business, for engineers, or for postgraduate study). The language needs of any student will be to some extent specific, even if his intention is no more than to use the language as a tourist. In all these cases, efforts must be made to use in the test the kind of language which actually occurs in the situations the student will meet. For example, a test which sets out to assess the proficiency of a student hoping to follow a university course in an English-speaking country would need to take into account not only his level of skills in listening to lectures, but also his ability to take notes, to make full use of what is written on the blackboard or projected, and finally to use the information he has gained from the lecture in his subsequent writing. An important element in proficiency testing is to assess in some way the student's ability to repair breakdowns in communication, by asking for a repetition or an explanation, for example, or by apologising and rephrasing what he has just tried to say.

This division of tests into separate categories is clearly rather arbitrary, and though it is convenient to say that the purpose of any test can be defined in this way, there are in practice several different purposes for every test. For example, it has been argued above that an achievement test assesses the learning that has gone before, and is therefore concerned with the past, but a student taking this kind of test usually does so because he needs the qualification to convince someone else of his future potential, either as a student in a more advanced course or as an employee.

In the same way, although a proficiency test is not in theory concerned with how the student's present stage of competence has been arrived at, he may well have prepared for it by taking a course designed to help him to pass.

The four types of test described above form the basis for the central part of this book, Chapters 4-7, in which techniques for setting tests are given in detail. Before these practical issues are reached, however, there follows a discussion of the fundamental qualities required in good language tests of all kinds.

2 Qualities of a good test

The three most important characteristics of a good test are reliability, validity and practicality. These abstract nouns may seem rather daunting, but just as it is impossible to play chess without knowing how a knight moves across the board, so it is pointless to try to write tests without a basic understanding of the principles behind them. A teacher who is unaware of the relationship between the content of a test and the consistency of its results is in danger of writing tests which produce misleading information about his students.

2.1 Reliability

The *reliability* of a test is its consistency. There would be little point in trying to measure people's waists with a piece of elastic. What is needed is a tape measure which stays the same length all the time, so that one person's waist is known to be eighty-one centimetres and another's ninety-one centimetres. Tests should not be elastic in their measurements: if a student takes a test at the beginning of a course and again at the end, any improvement in his score should be the result of differences in his skills and not inaccuracies in the test. In the same way, it is important that the student's score should be the same (or as nearly as possible the same) whether he takes one version of a test or another (for waist measurements, the same result should be obtained whichever tape measure is used), and whether one person marks the test or