

APPROACHES TO COMMUNICATIVE TESTING¹

Jeremy Harmer
Instituto Anglo-Mexicano
Guadalajara

A - INTRODUCTION

This paper is an attempt to take a brief look at a topic that has interested and worried test designers since the advent of more communicative approaches to EFL methodology. 'Communicative Testing' is often talked about but seldom put into practice. With few exceptions tests still resemble the approach so ably argued many years ago by Robert Lado (see below). And yet if the methodological implications of communicative syllabus design are fully adopted by the EFL profession, communicative test types must automatically follow if we are to have any reliable measurement of student progress. This paper will discuss some of the issues surrounding the concept of communicative testing and then look at four oral test types and suggestions for scoring.

B - TESTING AND ITS IMPORTANCE

This paper is particularly concerned with achievement testing: in other words I shall be looking at test types that are designed to elicit information about whether students have or have not achieved what was expected of them after a given period of instruction. Achievement testing is particularly vital for three main reasons:

- a) If the concept of testing has any validity it is that a test should provide the tester with valuable information. In other words, the teacher marking a class test should learn from it many things. He should learn both about the progress of individual students (i.e. 'how well has Pedro achieved the goals set for this semester/year etc.?') and about that student's progress within a group. A good test, in other words, should give the teacher information about students relative to each other. Such a test will cause the best students in the class to do better than weaker pupils. Lado (1961) referred to this as

¹ Some of the contents of this paper were first given at the Guadalajara and Puebla mini-conventions in 1979.

Test Reliability, meaning that a good test will reliably measure both a student's own progress, and his progress in relation to the class.

- b) The results of a good test will be invaluable for the teacher. He will get from those results not only information about his students, but also about the effectiveness of his teaching. Such information is, of course, vital to any teacher.
- c) Perhaps most importantly tests have a washback effect on what goes on in the classroom. Particularly where tests are mass produced for (e.g.) a state system, the design of the tests will dramatically influence the teaching that takes place in the classroom. Most teachers want their students to do well, both out of a real interest in their students and out of a fear of teacher failure, which all members of the profession suffer from to a greater or lesser degree. Where institutionalized tests have an unchanging format over a couple of years or more, teachers almost automatically find themselves adapting their methodology so that 'exam preparation' has more than its equal share of influence. Classic examples of this abound, notably in British EFL where the Cambridge exams have almost completely stifled any attempt to really investigate the needs of upper intermediate and advanced students. Upper intermediate students are invariably put through a rigorous programme of exam preparation for the Cambridge First Certificate exam regardless of whether that is the kind of learning which is appropriate to such a student. And of course the same goes on wherever a test or an exam is of importance, whether it is the first year of secondary school or a highly advanced group at a private institute. Tests and exams, in other words, have an importance far beyond the measurement of student ability. Their effect 'washes back' over the classroom, and methodological principles are wantonly sacrificed for success. And yet tests are frequently written at the last moment by underqualified people, little realising the effect of their actions. When we speak of communicative testing it is with a real sense of urgency since the gains of communicative syllabus design are in grave danger of being lost unless test instruments are designed which will, in effect, encourage the real practice of such methodology through the washback effect mentioned above.

For all these reasons the design of tests is of paramount importance. The test writer should be an expert, not just 'one of the teachers' since he wields enormous power, and since the design of effective tests is a highly skilled operation.

C - SCIENTIFIC TESTING AND LANGUAGE USE

The composition has often been used to measure language ability: it is a test type that has been used, literally, for years. The problem arises over the scoring, or marking, of such a testing device. It has often been suggested that were the same student composition to be given to five different scorers, five different scores would emerge. The scorer would naturally use his subjective judgement as to the merit of the composition, and each person's subjective judgement would differ. Hence the five different marks. This situation seems inequitable particularly where students are following parallel courses but have different scorers. Both may produce much the same work, but they may be scored differently and, of course, the same situation applies to oral tests where the examiner and the student engage in 'free' conversation. Not only is there injustice in such a situation, but also there is a basic lack of reliability. If the scoring of a test is variable in this way, how can we be sure that the results of the tests give us accurate information of the kind we have suggested we need (see b above)?

It was largely for reasons such as these that Lado (op. cit.) developed the scientific approach to testing that still has tremendous influence today. Lado was concerned with how to 'measure' a student's ability in a language so that reliable results were obtained, results that would adequately quantify this ability. This, of course, meant that he was concerned about the reliability of scorers for test types such as compositions. These two points were covered by his argument that tests should contain discrete items and be designed in such a way that scoring was reliable and objective.

The concept of a discrete item is simple: it is an item that stands on its own. In test terms this means that you test only one thing at a time. Thus an English test might have as its first item a multiple choice question to find out if the student knows the difference between 'some' and 'any'. The second item might test the student's ability to use the 's' for the present simple 3rd person singular etc. A test with a total score of 100, in other words, would contain 100 discrete items. The scorer or teacher is then in a position to say that student A, who scored 65, knows more English than student B, who scored 60.

Clearly this approach to testing is extremely attractive since the test designer is here in a position to decide exactly what it is he wants to include. He can be absolutely precise about the language that the students have to produce or recognise, and be sure that any given item is testing only and exactly what he wants to be tested.

Another great advantage of the discrete item test is that it is potentially much easier to score. Either the student got item 5 right or he didn't. This is a very different situation, of course, from the composition or similar type of item. And it goes a long way towards the ideal of scorer reliability. Tests should be designed where anybody scoring the test would arrive at the same result, and this is naturally easier to do with a discrete point test; particularly if it is in the form of multiple choice where the scorer's judgement is not a factor and where the only task he has to perform is marking the correct letter a, b or c. The ideal has been reached since a multiple choice test can be marked by anybody (whether or not they have a knowledge of English) or even by a computer. The scorer is completely objective and the scoring is completely reliable.

Certain worries about this type of test design do intrude, however. While the ideal of 'scientific' testing has been reached (e.g. reliability and objectivity), the ideal of finding out about a student's ability to USE the language has not. It would seem ridiculous to say that student A knows 65% of English because he got 65 out of 100 discrete items correct. The ability to use language is rather different from the ability to recognise or even produce individual grammar points. The student who knows the difference between the present simple and the present progressive does not necessarily know when it is appropriate to use them. And while it is true that the discrete point test may be appropriate for finding out if a student has memorised 100 chemical formulae (although this is doubtful), the same cannot be said for a skill like language, where the ability to use the language is as important as the theoretical knowledge of that language.

What is needed, then, is a test that will measure a student's ability to use the language, and this will clearly involve more than just an ability to discriminate between grammatical forms or select the correct form of a verb. A language user is forced to integrate a number of different skills and language exponents, and for this reason it seems vital for our tests to be integrative. In other words, we will test the student's ability to 'pull together' the various elements of his knowledge, and in this light the composition, or indeed any device to make the student produce a piece of connected discourse seems to be, after all, not such a bad idea. A student involved in a piece of connected writing has to select not only what grammatical forms are acceptable, but also what language is appropriate. He has to draw from all the language knowledge he has in order to do this. And the same is true, of course, for oral testing. A language user involved in a spoken interaction not only has to be able to select acceptable and appropriate language, but also has to be able to process what the other participant in the interaction is saying. Both a decoding skill (listening and understanding) and an encoding skill (selecting acceptable and appropriate language forms) are necessary, and the only real test of a

student's oral ability is to see how well he performs in such an interaction. It will be clear that a test composed entirely of discrete items will not give the scorer such information. Since only one thing is being tested at a time, we can hardly suggest that language use (as we have suggested, an integrated set of skills) is being measured.

The use of integrative test types is particularly important in the modern climate of communicative methodology. The whole emphasis in EFL during the past few years has been on teaching students how to use the language. Techniques such as group work and role play have come to prominence precisely because they enable the teacher to make some pretence at simulating the real world inside the classroom so that language practice can mimic reality. Grammar drills and grammatical knowledge are seen only as enabling steps, and not as end products in themselves. Both in speaking and writing students are encouraged to produce or interact in stretches of connected discourse as they would have to do in real life. Such methodological principles, though, are automatically undermined by the unwillingness of test designers to abandon the discrete point test, producing the washback effect we noted in Section B in this paper.

The problem of objective scoring still remains, however, and integrative test types will have to be designed so that the actual measurement of a student's ability can take place in the most effective way.

D - ORAL TEST ITEM TYPES

In this section we will look at four examples of oral test item types that aim to meet the ideal of communicative testing to a greater or lesser degree. The problems of scoring will be left until section E of this paper.

While it is true that 'conversation' has a lot to recommend it as a testing device, it is felt that communicative testing of oral skills, particularly, should in general be task-oriented. Earlier we said that the best test of a student's ability was to see how well he managed in an oral interaction. Clearly one of the measures of such ability is whether the task he had set himself (or had had set) was achieved (e.g. finding out how to get to the station, conveying disapproval, etc.). The test item types below, therefore, are based on the student's ability to achieve some sort of a task.

1.- INFORMATIONAL TESTS.

Here both an oral examiner and the student have certain pieces of information, and the test will center around that information.

a) Question answering.

Here the student has certain information in front of him, and he will be asked about that information by the examiner.

Example: Student's copy (a railway timetable)

	Cost	Frequency
London-Birmingham	£ 5.00	Every 30 mins.
London-Leamington Spa	£ 4.30	Every 40 mins.

The examiner will ask the student such questions as:

How often do trains go from London to Leamington Spa?

Clearly this type of test item is limited, but it does give the examiner an opportunity to assess a student's question answering (and understanding) ability. The student's task is to be able to give answers to questions about information he knows - a natural function for most language users.

b) Question Asking.

Here both the student and the examiner will have information, but the student's copy will be missing certain information. The student's task will be to complete that information by questioning the examiner.

Example: Student's copy.

Examiner's copy.

HOUSE FOR SALE:
Rooms: 1 Kitchen, 2, 3,
<u>Additional features</u>
<u>Cost</u> :
<u>Location</u> :

HOUSE FOR SALE:
Rooms: 1 Kitchen, 2 bathrooms, 3 bedrooms, 1 dining-room, 1 living room.
<u>Additional features</u> : Garage, garden.
Cost \$20,000.00
Location: Country, 6 miles from city centre.

a) Question answering.

Here the student has certain information in front of him, and he will be asked about that information by the examiner.

Example: Student's copy (a railway timetable)

	Cost	Frequency
London-Birmingham	£ 5.00	Every 30 mins.
London-Leamington Spa	£ 4.30p	Every 40 mins.

The examiner will ask the student such questions as:

How often do trains go from London to Leamington Spa?

Clearly this type of test item is limited, but it does give the examiner an opportunity to assess a student's question answering (and understanding) ability. The student's task is to be able to give answers to questions about information he knows - a natural function for most language users.

b) Question Asking.

Here both the student and the examiner will have information, but the student's copy will be missing certain information. The student's task will be to complete that information by questioning the examiner.

Example: Student's copy.

Examiner's copy.

<p>HOUSE FOR SALE:</p> <p>Rooms: 1 Kitchen, 2, 3,</p> <p><u>Additional features</u></p> <p><u>Cost:</u></p> <p><u>Location:</u></p>

<p>HOUSE FOR SALE:</p> <p>Rooms: 1 Kitchen 2 bathrooms 3 bedrooms, 1 dining-room, 1 living room.</p> <p><u>Additional features:</u> Garage, garden.</p> <p>Cost \$20,000.00</p> <p>Location: Country, 6 miles from city centre.</p>

The student will be told that he has to fill in as much information as he can about the house for sale on the basis of information he gets from the examiner.

Despite the apparent complexity of this task, it would seem to be a reasonable attempt at a communicative testing device.

The ability to find out information and process it is absolutely vital to any language user.

2.- ROLE PLAYING

We will look at two possible role-playing items, although there are obviously many more possibilities.

a) Controlled Conversation.

Both student and examiner are given half a conversation which they have to 'act out'. The examiner's copy, though, will have the actual words, the student's only a summary.

Student's copy

You are B and should follow the instructions you are given. You are talking to your friend A.

A speaks first.

A: _____

B: Refuse politely.

A: _____

B: Agree reluctantly, but ask A to return it.

A: _____

Examiner's copy.

A: Could you lend me your watch?

B: Refuse politely.

A: Oh dear! Couldn't I borrow it for just a bit? Mine's broken and I need to know the time.

B: Agree reluctantly, but ask A to return it.

A: That's awfully kind of you.

The level at which this conversation is 'guided' clearly means that it will not really be like authentic language use. Nevertheless, a student who performs well in this type of test would presumably have a very good chance of communicating efficiently in real life. It is interesting to note the difference between this type of test item, and the traditional 'situational response' which it in some ways resembles. The difference, of course, is that here some kind of interaction is taking place.

b) Free Conversation.

Here the student is told to 'act out' a role based on certain information.

Example: Student's copy.

You are in a restaurant.
 You are hungry and thirsty,
 You have only £ 4.50
 You are talking to the waiter.

Examiner's copy.

You are a waiter. There is no written menu, but you are prepared to give the student the following information:

Beer: (£0.75) per glass.
 Soup: Onion (£0.95) Tomato (£0.80).
 Starters: Avocado Vinaigrette (£1.50). Melon (£0.90)
 Prawn cocktail (£ 1.75)
 Main Courses: Chicken (£ 2.50) Roast Beef (£ 2.75)
 Liver (£ 2.00)
 Sweets: Flan (£ 0.70) Banana Split (£ 1.25)

Here the student has been put in a 'real life' situation and has to negotiate his order as he would have to do in a real restaurant. The achievement of his task (i.e. ordering a meal that he can afford) will prove a certain level of communicative efficiency.

Role-playing as a testing device has often been criticised on the grounds that some students find it very unnatural and difficult. The same, however, can be said of tests in general, and certainly about (e.g.) the ability to order a meal in a foreign country!

These four examples show how a student can be made to interact with the examiner in such a way that some semblance of real communication takes place. The tasks are as realistic as possible in the circumstances, and the examiner does not have to search his mind frantically for conversation initiators. With the type of scoring shown below, measurement can be more objective and reliable.

E - SCORING ORAL TESTS

We have said that our test item types involve the student in some kind of task performance, and clearly a measure of the student's ability will be the degree to which he achieves the task. But it is also true that efficient communication can take place even where linguistic ability is poor. In other words, a student might achieve the task even though his level of English was very low. Sidestepping, for the moment, the question of whether such a student could be considered communicatively 'competent', we will also want a measure of the student's linguistic ability. The proposed scoring system, therefore, will be divided into two parts.

1.- TASK ACHIEVEMENT.

Here the student will be marked on a scale of (for example) 0-5, where 5 is awarded to a student who successfully completes the task and 0 is awarded to a student who completely fails to achieve the task. In the second examples of both informational tests and role-playing (see above) this will obviously present no problem; but in the first examples of both the categories, task achievement will be based on whether the student gives the required information. In 1a this will involve the ability to give prices and frequency and in 2a whether the student can show that he is, for example, refusing.

2.- LINGUISTIC ABILITY

Linguistic ability will also be marked on a scale but this time it will be considerably more complicated since we will be interested in various areas of the student's ability. The following example, adapted from Valette (1977) measures students' fluency, vocabulary, grammar and pronunciation.

FLUENCY:	Speech is natural and continuous	4
	Speech is generally natural and continuous, but not always.	3
	Some definite stumbling, but manages to rephrase and continue	2
	Long pauses, utterances left unfinished	1
VOCABULARY:	Consistent use of appropriate words	4
	Minor lexical problems, but generally appropriate	3
	Vocabulary often inappropriate	2
	Vocabulary inappropriate and inaccurate	1
GRAMMAR:	No errors of morphology/syntax	4
	Occasional slight errors, nothing serious	3
	Occasional serious errors + more slight errors	2
	Serious errors throughout	1
PRONUNCIATION:	Phonemically accurate throughout	4
	Occasional pronunciation error, but always comprehensible	3
	Pronunciation errors causing difficulty in perceiving meaning.	2
	Pronunciation problems cause incomprehensibility	1

With these two measures we will be able to say something about the student's communicative efficiency and about his linguistic ability. Some elements of scorer subjectivity will still exist, but they have been considerably reduced.

Marking scales of this type are not only appropriate for oral tests, of course. The same principle can be applied to writing where marks may be awarded for such things as task efficiency, grammar, vocabulary, punctuation, style, etc.

F - CONCLUSION

Communicative testing is still in its infancy even though communicative methodology is widely in use in the classroom. For many reasons it will rely largely on integrative testing devices while some attempt at scorer reliability will be made. As we have said, the need for more and better communicative tests is pressing.

I have tried to show four examples of test types that are 'communicative' in the sense that they involve some kind of realistic interaction, and I have tried to suggest that the scoring of such tests need not be as subjective as some assume.

BIBLIOGRAPHY

A particularly useful document from which some of the ideas in this paper have been taken is:

Morrow (1977) Techniques of evaluation for a Notional Syllabus.
Study commissioned by the Royal Society of Arts.

The two works referred to in this paper are:

Lado (1961) Language Testing. Longmans.

Valette (1977) Modern Language Testing. Harcourt Brace Jovanovich.