

The Testing Problem : Communicative Tests for Communicative Lessons

Stephen M. Ryan

Department of English

Mukogawa Women's University, Nishinomiya 663

The idea of Communicative Language Testing is a necessary adjunct to the concept which has been developed since the late 1970's of a communicative approach to language teaching.¹ Since the aim of the latter is to develop students' communicative ability in the target language, it follows that any form of test given to these students should be designed to measure and chart the progress of such communicative ability.

It has proved, however, extraordinarily difficult to develop tests that measure this ability and nothing else.² Short of putting the students in a target-language community for some time and grading their ability to cope with day-to-day situations as they arise, how can we ever judge their ability to use the language for effective communication in realistic situations?

Answers to this question can involve the need for huge resources, not least of time and imagination. Since such resources exist only in a very few cases, it seems wise to begin from the actual situation where a test is required, in order to keep the solution to the testing problem within the bounds of practicality. Consequently, this paper will begin by describing the situation in which the author finds himself, continue with a series of theoretical assumptions about testing and show the constraints these practical and theoretical considerations impose, before presenting a possible solution to the problem.

The Situation.

I teach a course called, "English Conversation," to classes of first year University and Junior College students. Some of the students come from high schools where they have been encouraged to speak and listen to English but there are so few such students that all are assumed to be starting from zero as far as English conversation is concerned. The aim of the course, then, is to enable the students to activate the passive knowledge of English they have gained from 6 years of studying English in Junior and Senior High School so that they are able to use the language to communicate in everyday situations. To achieve this aim, great emphasis is placed on students' using the language they have in an unselfconscious way to achieve specific communicative goals, such as asking the way, talking about their families and making plans for the future.³

Each class consists of approximately 30 students. I have 8 classes in all, comprising over 240 students. Each class meets twice a week for 45 minutes each time.

The university administration requires that each student be given a percentage grade at the end of each semester. Time is set aside each semester for students to take written examinations. The pass mark is 50% and students awarded less than that are allowed to take a make-up test.

The Assumptions.

The following assumptions are made about testing :

- 1) that the main purpose of testing is to make each individual student aware of her strengths and weaknesses.
- 2) that the method and content of the test should be consonant with the aims of the course. (Content validity.)
- 3) that the method and content of the test have a strong influence on students' attitudes to course-work. (Back-wash.)
- 4) that grades should not be based on entirely subjective criteria. (Objectivity.)
- 5) that the student should feel that the test is a true measure of her abilities. (Face validity.)

1) The Purpose of Tests.

This assumption informs all the others. It sees tests not as an enemy for students to fight in order to gain a certain percentage or ranking in the class but as an ally that will help them to know where they are doing well and where they need further effort. Such tests perform a similar function for the teacher in revealing the strengths and weaknesses of individual students and the class as a whole.

Testing, then, is seen as a useful adjunct to teaching, not as its purpose.

2) Content Validity.

This assumption rests primarily on the principle that it is unfair and uninformative to test something that has not been taught. This means that knowledge not given and skills not practised should not be tested. It also means that the kind of test given should be appropriate to the lesson-content: a reading test is inappropriate for a listening course, a written test for a speaking course.

3) Back-wash.

The basis for this assumption is apparent whenever teachers talk about testing. At its most basic level, it says: "If I don't give them a test, they won't learn it," and, by extension, "They won't take the course seriously unless there's a test at the end of it."

That the same principle applies on the level of content is apparent from the notorious example of Japanese University entrance examinations. The back-wash from these important tests is so strong that they dictate the whole of the high school syllabus. Students do not want to, "waste their time," studying anything that will not be tested.

From the students' point of view the test is a message from the teacher to tell them which parts of the classwork are important and which can be ignored as they will not be tested.

4) Objectivity.

The main advantage of objectivity is that it allows students to see exactly what is required of them (the criteria for full marks) and, after the test, to see in what ways they have failed to reach the standard. Being able to present an objective series of criteria on which the grade is based also adds to the face validity of the test.

5) Face Validity.

In order for the test-grade to have meaning for a student, she must feel that the test was a fair one. If it is not so perceived a bad mark will be blamed on an unfair testing procedure and fail to have the desired result of exposing weaknesses on which the student should focus her energies.

The Constraints.

These assumptions interact with the situation in which I am working to produce a series of constraints on the contents and procedure of any test that is to be given.

Content Validity.

Any written test would lack content validity for this course : the course teaches conversational skills that cannot fully be tested on paper. To test only those sub-skills that can be reproduced in writing would give them undue emphasis and is likely to have a distorting effect on students' perception of English Conversation because of the kind of back-wash such a test would create.

Among the different kinds of speaking tests available, the principle of content validity is also a constraining factor. Asking the student to give a short speech or interviewing her⁴ are unacceptable as speeches and interviews are distinct modes of discourse that have little place in a conversation course.

Time.

Since a written test is inappropriate, tests must be held during class time. This means that every minute spent on testing means one minute less spent on coursework. It thus becomes important to see that this time is well spent.

The kind of test in which students come one by one to speak to the examiner⁵ are time-consuming and must therefore be of undoubted usefulness if they are to be considered.

Objectivity.

If grading is to have an objective basis it is not enough for the teacher to listen to the students conversing (whether with each other or with the examiner, in person or on tape) and announce a score based on general impression or even on his impression of their performance in particular categories (Grammatical Accuracy, Pronunciation, Communicative Ability, etc.).⁶ Some quantifiable and justifiable measure must be found.

Consistency.

If the students do not perceive the test to be the same for all of them it will lose its face validity. Unstructured conversation, then, is not acceptable as a test since student A's conversation may involve language far more complicated than Student B's. Some kind of reproducible structure is needed.

Affect.

The principle of face validity requires that students' attention be focussed on what is actually being tested, otherwise they will complain that the test was unfair. If, for example, the test is highly stressful, students will focus so much attention on their own nervousness that they will be able to say, quite rightly, "I was too nervous to do my best." Stressful situations should therefore be avoided as far as possible.

The Test.

Faced with the constraints described above, I have been evolving a suitable form of test. What follows is not a definitive solution to the problem but the latest in a series of steps towards that goal.

Students are all tested at the same time. They sit in pairs, each pair as far from the others as possible. Each student faces her partner across two desks. In the middle of the desks is a bag which prevents each student from seeing the other's test paper. Students are forbidden (on pain of failing the test) to speak Japanese or look at each other's paper.

The paper instructs the students to carry out some communicative task together. They may have to plan a party, share information about their families, describe the layout of a room or ask each other a series of questions.⁷ Whenever a student receives information from her partner (hopefully through the use of English, although gesture is not outlawed) she is to record it quickly on her paper, either as a picture or a verbal note, eg. "Born—June 2nd."

The object is not to complete the task but to perform as much of it as possible within the allotted time (usually 5 minutes).

The teacher does not listen in as they perform the task. With a room full of 15 pairs this is not possible. He simply watches and listens to see that no cheating is taking place.

Assessment is based entirely on the amount of information exchanged or, to be more precise, the amount of information recorded on the test paper. This criterion rests on the assumption that the more communicative competence a student has in English the more efficiently she will communicate and thus the more information will be conveyed in the limited time. Each student's score is added to that of her partner. It is this combined score that is then compared with those of other pairs and expressed as a percentage of the score of the highest scoring pair. Combined scores are used because it is thought to be impossible to disentangle who is responsible for the successful recording of each piece of information: its initiator or its recorder.

Each student takes three such tests, performing three different tasks with three different, randomly selected partners. This produces three percentage scores for each student. The highest one is rejected on the assumption that it is mainly the work of a good partner. The lowest one is also rejected for similar reasons, leaving the middle score as the student's exam grade.

Remarks.

The test described above can be seen to comply with the constraints imposed by the situation and the assumptions about testing in the following ways:

Content Validity.

Students are performing exactly the kinds of activities they have been practising during the class-time. The only differences are: cheating is more heavily punished, partners are randomly selected and the students cannot call on the teacher for help. There have been no complaints from the students that the test was unfair in the sense of not testing the kind of English they had practised in class.

Time.

All three tests fit easily into one 45 minute lesson. Up to 8 minutes speaking time can be allowed per test and still leave time for the admittedly cumbersome logistics of repeatedly rearranging desks, test papers and students.

Objectivity.

Test results are objectively based in that they correspond directly to the number of pieces of information that the grader finds on the papers of both members of a pair. There are at least three possible objections to this:

- a) that more communication may occur than is recorded on the paper.

In order to impress on the students the importance of recording each piece of information they exchange, a rehearsal for the test, involving a task that will not be part of the real test, is carried out beforehand. At this time the marking system is thoroughly explained.

- b) that the combined total may reflect the efforts of one partner more than those of the other.

It is for this reason that three tests are conducted rather than one.

- c) that the grader may not always be able to distinguish what constitutes, "one piece of information."

Care is taken to ensure consistency in this respect.

While none of these countermeasures is thought of as overcoming the barriers to objectivity, they do ensure that grading is not a completely subjective process.

Consistency.

The structure provided by the instructions on the test paper is exactly the same for each pair of students. The time allowed is also demonstrably the same. In this sense consistency is achieved.

The great problem of consistency is that each student's score depends to a certain extent on the

linguistic abilities of her partner. Clearly, the greater the number of different tests taken with different partners the less important this factor becomes. In this respect, the practice of each student taking three tests with three different partners is a compromise between the constraints of time and consistency.

Affect.

Students new to the test comment again and again on how relaxing it is to take the test with a classmate, without a teacher listening to and judging every word. A certain amount of tension is inevitable and perhaps desirable in a test but students say that this element is nowhere near as bad as they had imagined on first hearing that they were to take a speaking test.

Some students experience unease because this kind of test is so unlike any they have taken in the past. This unease can easily grow into mistrust of the test itself. It is hoped that the rehearsal and explanation of the test will reassure these students to a certain extent.

Conclusion.

As the foregoing remarks suggest, this test is far from being an ideal solution to the testing problem. However, it is an approach that seeks to fit in with both the practical and theoretical constraints on testing in this particular situation.

It is heartening that the greatest number of complaints from students dissatisfied with the test concern the fact that it does not test memorized knowledge. Heartening because this is precisely the point: communicative competence in English, which this course attempts to teach and test, is a skill rather than a simple matter of accumulated knowledge.

Notes.

- 1 See especially, Brumfit, Christopher and Keith Johnson (ed) *The Communicative Approach to Language Teaching*. Oxford : OUP, 1979.
- 2 See, amongst others, Underhill, Nic. *Testing Spoken Language*. Cambridge : CUP, 1987.
- 3 For details of how this is done, see my, "Language Fluency Practice in an Eraser Culture," in *Mukogawa Kyo*, No. 36 (1988). 11-17.
- 4 Both these procedures are used by the Society for Testing English Proficiency in their, "Eiken," oral tests.
- 5 This procedure has been used by local examination boards for foreign language "O" and "A" level exams. in England and Wales.
- 6 These categories are used in just such a way in the University of Cambridge Local Examination Syndicate's "Cambridge Eiken," English language oral tests.
- 7 For a specimen pair of exam. papers see Appendix.

(1989年9月27日受理)

Student 1

Ask your partner about the times of the following things and write them below :

The Play	_____	Macy's	_____ - _____
The Movie	_____	Citibank	_____ - _____
The Ballet	_____	The Post Office	_____ - _____
The Concert	_____	Rexall Drugs	_____ - _____
The Opera	_____	The Museum	_____ - _____

Use this information to answer your friend's questions :

Airline	Flt.	Destination	Dep.	Arr.
Air France	AF942	Paris	7.30 P.M.	8.35 P.M.
Vang	VA358	Rio	11.25 P.M.	9.25 A.M.
Japan Air Lines	JAL27	Tokyo	12.40 P.M.	5.20 P.M.
Lufthansa	LH635	Frankfurt	8.25 P.M.	9.50 A.M.
PanAm	PA726	Amsterdam	7.10 P.M.	10.15 A.M.
Qantas	Q411	Sydney	9.00 P.M.	6.05 A.M.

The map shows the following facilities and their hours:

- Medical Center**: Monday-Friday 10 - 5 evenings 6 - 9
- City Art Gallery**: Tuesday-Friday 10 - 6, Sat and Sunday 10 - 5
- Public Library**: Monday-Friday 10 - 6
- Bank of America**: Hours Mon - Friday 9 - 3
- Central Department Store**: Monday-Friday 10 - 10, Sat-Sun 12-5

Your name _____
Your partner's name _____

Student 2

Ask your partner about the times of the following things and write them below :

The Library	_____ - _____	Flights to:	
The Art Gallery	_____ - _____	Amsterdam	_____ - _____
The Bank	_____ - _____	Paris	_____ - _____
Central Department Store	_____ - _____	Tokyo	_____ - _____
The Medical Centre	_____ - _____	Sydney	_____ - _____
		Rio	_____ - _____

Use this information to answer your friend's questions :

Post Office	9:00 - 5:00
Citibank	9:00 - 3:00
Museum of Modern Art	11:00 - 6:00
Macy's	9:00 - 6.45
Rexall Drugs	8:00 - 7:30

ENTERTAINMENT GUIDE

OPERA AND BALLET
"The Nutcracker"
American Ballet Theater
7:30 tonight

CARMEN
tonight
at the Metropolitan Opera House 7:30

THEATER
BEST PLAY OF THE YEAR!
8:00 p.m.

CINEMA
TOOTSIE
Don't miss this movie!
Dustin Hoffman at his best.
7:20

Philadelphia Symphony Orchestra this week at LINCOLN CENTER all Beethoven program 8:00

Your name _____
Your partner's name _____