

# Coming to Grips with Progress Testing: Some Guidelines for Its Design

BY CARMEN PÉREZ BASANTA

This article was first published in Volume 33, No. 3 (1995).

The area of progress testing has been neglected and has lagged far behind developments in language teaching and testing in general. In most classrooms today, English is taught through communicative textbooks that provide neither accompanying tests nor any guidance for test construction. Teachers are on their own in constructing tests to measure student progress and performance. The result is they write traditional grammar-based items in a discrete-point format that does not fit the communicative orientation of the textbook or the underlying teaching principles.

In many cases, teachers have been reluctant to administer regular tests. Stevenson and Riewe (1986) give the following reasons for this:

1. Teachers consider testing too time-consuming, taking away valuable class time.
2. They identify testing with mathematics and statistics.
3. They may think testing goes against humanistic approaches to teaching.
4. They have gotten little guidance in constructing tests in either pre-service or in-service training.  
Personally, I would add:
5. Teachers feel that the time and effort they put into writing and correcting tests is not acknowledged with additional pay or personal praise.
6. There is the personal implication that I would call “the image in the mirror”: Testing puts you face-to-face with your own effectiveness as a teacher. In this sense, testing can be as frightening and frustrating to the teacher as it is for the students.

## Why Must Teachers Test?

If we assume that a well-planned course should measure the extent to which students have fulfilled course objectives, then progress tests are a central part of the learning process. Other reasons for testing can be identified:

1. Testing tells teachers what students can or cannot do—in other words, tests show teachers how successful their teaching has been. It provides washback for them to adjust and change course content and teaching styles where necessary.
2. Testing tells students how well they are progressing. This may stimulate them to take learning more seriously.
3. By identifying students’ strengths and weaknesses, testing can help identify areas for remedial work.
4. Testing will help evaluate the effectiveness of the programme, coursebooks, materials, and methods.

This continuous feedback provided by tests will benefit students, who will feel that their weaknesses are being properly diagnosed, and their needs met.

## Theoretical Considerations

As the majority of teachers have not received enough training in test development, let me suggest a framework for the design of tests that fit with classroom activities. Let us start by defining progress tests as a measure of students’ progress towards definite goals. In this sense we do not make any distinction between progress or achievement tests: we conceive of both as one means for monitoring performance and evaluating the final outcome.

The second important issue is whether there is a discrepancy between teaching and testing. Weir (1990:14) has pointed out that the only difference between teaching and testing within the communicative paradigm relates to the amount of help that is available to the student from the teacher or his/her peers. Still there are some constraints that the process of testing imposes, such as time, anxiety, grading, and competition. But, on the whole, we agree with Davies (1968:5) when he says that a good test is an obedient servant since it follows and apes teaching. Our tests should be based on the classroom experience in terms of syllabus, activities and criteria of assessment. Their final aim is to measure the language that students have learned or acquired in the classroom both receptively and productively. We could conclude by saying that the more our tests resemble the classroom, the more valid they will be.

The theoretical requisites that a test must achieve are validity, reliability, and practicality.

A test is valid if it measures what you want it to measure.

Construct validity refers to the concomitance between the test and the underlying teaching principles. It follows from this that tests should reflect the objectives of the course and underlie its teaching principles. As regards communicative testing, it is crucial that tests be as direct and authentic as possible; they should relate to real life and real communicative tasks.

A progress test has content validity if it measures the contents of the syllabus and the skills specified in the coursebook. Hence, we should take into consideration the learners' needs and their particular domain of use to ensure content validity. Success with regard to this aspect is quite easy to achieve since the coursebook designer has decided on the course content. The task of the test writer—the teacher—is to sample this domain, measure it, score it, set up pass/fail cutoffs, and give grades.

If a test is appealing to laymen—students, administrators, etc.—it has face validity. In other words, tests should be based on the contents of the textbook and the methodological teaching approaches, as well as measuring what it is supposed to measure.

Tests are reliable if their results are consistent, i.e., if administered to the same students on another occasion, they would obtain the same results. There are two main sources of reliability: the consistency of performance from candidates and scoring.

Finally, a test has practicality if it does not involve much time or money in their construction, implementation, and scoring.

## Planning Stage

*Specifications.* Even if the specifications were done by the textbook writer, the teacher will have to select what s/he considers most important, and not what is easiest to test, in order to draw up a set of specifications which reflects the emphasis of the teaching (McGrath and Kennedy, 1979). Thus, in this stage, we aim at ensuring content validity which, as Anastasi (1982:131) defines it, is “essentially the systematic examination of the test content to determine whether it covers a representative sample of the behavioural domain to be measured.”

As far as construct validity is concerned, there are certain features of communicative language teaching that we should attain within the testing format: demand for context, information gap, unpredictability, authentic language, participant roles, emphasis on the message, integration of skills, emphasis on discourse, and real life situations.

Two main implications may be drawn from these principles. The first is that we will have to concentrate both on use and usage. The second involves a reconsideration of the authenticity of texts and tasks. Authentic texts are not problematic but the fact that tasks should be based on real life contexts may present difficulties. As Picket (1984:7) puts it: “By being a test, it is a special and formalised event distanced from real life and structured for a particular purpose. By definition, it cannot be real life that it is probing.” In the same sense, Alderson (1981:57) states that the “pursuit of authenticity in our language test is the pursuit of a chimera.”

But communicative testing is as communicative or non-communicative as communicative teaching, in so far as directness and authenticity of performance are always restricted under classroom conditions. But, even if we admit that real life, authentic situations are not fully attainable, we should aim not to test how much of the language someone knows, but his ability to operate in a specified sociolinguistic situation with specified ease or effect (Spolski, 1968:92).

*Sampling.* Tests should cover the language, grammar, vocabulary, phonology, functions, and skill areas. Therefore, they have to cover both the content input and the activities or tasks. A test of communicative competence should test usage as well as the ability to use the language appropriately. If we want testing to accord with teaching, there should be a complete harmony between our teaching and our testing specifications. We will test what we teach and in the same proportions.

## Development Stage

In this stage we start the process of test design. I propose the following guidelines for their construction:

1. Compile written and spoken source materials that fit the contents of the programme. As Carroll and Hall (1985:18) have stated, these inputs should be authentic, coherent, comprehensible, at a suitable level of difficulty, and of interest to learners. These materials can be obtained from newspapers, advertisements, leaflets, stories, etc. It is useful to group them under different themes and to identify the proficiency levels for which they are appropriate.
2. Select activities that best measure performance. We should try to include all the possible activities used in the classroom.
3. Select test format—multiple choice, true/false, gap filling, etc.—taking into account channels, written or spoken, and strategy use.

The selection of test format is fundamental and controversial. Carroll and Hall (1985) classify them into three categories: a) Closed-ended, b) Open-ended, and c) Restricted response. The first category is analytical and objective and should be used for the receptive skills of reading and listening. The second category, manifested in essay/composition tests and interviews, is subjective, impressionistic, and global. The third category is content-controlled but may allow for more than one answer.

4. Avoid items that are ambiguous, tricky, or overlapping. The difficulty should lie in the text and not in the question. For every item, teachers should be able to identify which strategy we want to tap into. All methods may be valid as long as they are well constructed, and their selection will depend on what is to be tested. The inclusion of as many methods as possible will palliate the negative effects of using just one.
5. Include clear and unambiguous instructions, with brief and well-chosen wording and some examples. Weir (1993:24) recommends instructions to be candidate-friendly, comprehensive, explicit, brief, simple, and accessible.
6. Design a clear layout which will not induce mistakes. Make the test attractive, and similar to the layout of the textbook. We recommend variety, such as the use of pictures, different typefaces, and any element which can reduce anxiety.
7. Thoughtfully consider the scoring and marking systems. Testing is a teamwork activity not a solitary one. The marking system should be checked by at least another teacher. The marking criteria should be set beforehand and candidates must be informed as how they will be scored.

There are two ways of marking: by counting and by judging (A. Pollit, 1990). The former is the objective pro-

cedure in which the answers are either correct or incorrect, mainly used for testing the receptive skills. The latter is subjective and used for the productive skills. One way of making subjective, impressionist judgements more objective is to devise a marking scheme through bands and scales in which the judging criteria is described as precisely as possible. These bands should be made as simple and intelligible as possible (e.g., fluency, range of vocabulary, accuracy, appropriateness, etc.) so that scorers will not have to take into account too many aspects at the same time.

8. Analyse the test statistically. Basic statistics are more straightforward than we imagine. Calculate the reliability coefficient—Kuder-Richardson—and the difficulty and discrimination coefficients. The first mathematical operation tells you how reliable a test is; the other two measures show if the items are at the right level of difficulty and how well they discriminate. These mathematical operations are simple enough to be carried out in a manual calculator, and they can indicate the validity of the test and the performance of the examinee.
9. Consider the pedagogical effects that the test may have on teaching. Morrow (1986) stated that the most important validity of a test was that which would measure how far the intended washback effect was actually realized.

If we want our test to influence teaching and learning, we should ask our students and ourselves the following questions:

- What do students think about the fairness of the test?
  - What poor results are due to poor item construction? How could the items be improved?
  - What poor results are due to poor or insufficient teaching?
  - What poor results are due to the coursebook or other materials?
  - What areas of weakness in student performance have we detected for remedial work?
  - Can we make any assumptions on the relation between teaching and learning?
  - What changes should be implemented in our classroom as a result of the test feedback?
10. Present the test and feedback results to the students with the aim of reviewing and revising the teaching of content or skills in which the test has shown students to be weak. Teachers should listen to what students have to say about the test and profit from their comments.

## Conclusion

Teaching and testing are two inseparable aspects of the teacher's task. In spite of the current reluctance to profit from the latter, this article contends that testing has an essential role in the development of students' communicative competence. The brief nature of the article does not allow for an exhaustive description of progress testing. My intention is to encourage teachers to read more on the subject and to try some of the suggestions given.

## References

- Alderson, J. C. 1981. Report of the discussion on communicative language testing. In *Issues in Language Testing*. ELT Docs, III. ed. J.C. Alderson, and A. Hughes London: The British Council.
- . 1990. Bands and scores. In *Language testing in the Nineties: The communicative legacy*, ed. J. C. Alderson and B. North. Oxford: Modern English Publications.
- Anastasi, A. 1982. *Psychological testing*. London: Macmillan.
- Carroll, B. and P. J. Hall. 1985. *Make your own tests: A practical guide to writing language performance tests*. New York: Pergamon.
- Davies, A. 1968. *Language testing symposium: A psycholinguistic perspective*. Oxford: Oxford University Press.
- Morrow, K. 1986. The evaluation of tests of communicative performance. In *Innovations in Language Testing*, ed. M. Portal. London: Nfer Nelson.
- Picket, D. 1984, cited by P. Dore. 1991. Authenticity in foreign language testing. In *Current Developments in Language Testing*, ed. S. Anivan. Singapore: SEAMO Anthology Series.
- Pollit, A. 1990. Giving students a sporting chance: Assessment by counting and by judging. In *Language Testing in the Nineties*, ed. J. C. Alderson and B. North. Oxford: Modern English Publications.
- Porter, D. 1990. Affective factors in language testing. In *Language Testing in the Nineties*, ed. J. C. Alderson and B. North. Oxford: Modern English Publications.
- Spolsky, B. 1968. Language testing: The problem of validation. *TESOL Quarterly*, 2, 2.
- Stevenson, D. K. and U. Riew. 1981. Teachers' attitudes towards language tests and testing. In *Occasional Papers, 29: Practice and problems in language testing*, ed. by T. Culhane, C. Klein-Braley, and D. K. Stevenson. Department of Language and Linguistics, University of Essex.
- Walter, C. and I. McGrath. 1979. Testing: What you need to know. In *Teacher Training*, ed. S. Holden. Oxford: Modern English Publications.
- Weir, C. 1988. *Communicative language testing, with special reference to English as a foreign language*. Exeter University: Exeter Linguistic Series, 1.
- . 1993. *Understanding and developing language tests*. Hemel Hemstead: Prentice-Hall International.