

PROGRESS TESTING - PREPARATION AND ANALYSIS

Ghulam Muhammad Solangi
Syed Iftekhhar Hussain Jaffri

One of the main task of any educational institution and its teachers is to check the success of their forces with reference to particular students and the whole group. In the case of an individual student, there is a need to know how well he is keeping up with the programme should there be difficulties, the needs to be provided with, any remedial programme or to be moved to a more appropriate course. Checking on group progress enables a check to be made on the course itself and can provide clues as to how it can be improved.

The importance of the *progress test* will vary from situation to situation. In the small situation, where there are close contacts between staff and students, the teacher's assessment of students will probably have much greater weight than in the large institution with many classes running at the same time or where circumstances impose a considerable number of teacher changes. In the later case, there will be much more pressure for a common yardstick by which to judge student's progress. However, even in the small school a double check, such as provided by a progress test, will help the teacher to review his assessment and will help to indicate what areas of the course have yet to be mastered. A good *progress test* can also help the student to understand where his weaknesses lie and build up confidence by indicating the areas he has mastered.

Many of the principles for preparing good progress tests are similar to those for preparing achievement and proficiency tests. Instructions to the students should be clear, so that they have no difficulty in understanding what they have to do. Tricky questions should be avoided.

They are more likely to trick the better students than the poor ones. However, in other ways, the principles will be quite different, reflecting the different purpose the test is to serve. J.B.Heaton's (1975) quotation will serve to summarise these differences:

"Good performances act as a means of encouraging the student, and although poor performances may act as an incentive to more work, the *progress test* is chiefly concerned with allowing the student to show what he has mastered. Scores on it should thus be high (Provided, of course, that progress had indeed been made). Whereas in standardised achievement and proficiency test a wide range of performance should be indicated. The *progress test* should show a cluster of scores around the top of the scale. ¹

This difference has important consequences when preparing and analysing *progress tests*, and the techniques used for standardised testing are not completely applicable. This has not always been sufficiently emphasised. Some suggestions are now offered which may help produce good *progress tests* as defined in the above quotation.

Psychological tests have been developed for such a wide variety of purposes that testing method vary greatly from test to test. A number of continuo exist along which individual test can be classified.

"Test content can measure maximum performance or typical performance, the best one can do versus how one typically performs. Maximum performance tests have correct or incorrect answers. Typical performance tests generally assess stylistic differences, without specific answer being uniformly better than others".²

PREPARATION

One of the requisites of a successful syllabus is a clear statement of aims and methods, which is also a

requisite for successful testing. As the progress test should reflect the syllabus, the statement of aims and methods prepared for the syllabus will largely dictate the form the progress test will take.

Each course syllabus, except those for an absolute beginner in a particular language, will assume an ability in certain areas of the language. On the basis of this previous ability, the course will aim to develop abilities in new areas. In the progress test that follows such a course, the design and content will seek to show that the students have attained those abilities the course sought to develop. The test may include questions that involve students using the ability they were assumed to have at the beginning of the course, but it should not be central to the rest. Nor should the test include anything which was not considered a necessary previous ability, and which was not covered in the course. In other words, the content of the test should be such as to result in any student who has successfully mastered the course content getting perfect or near perfect scores.

Once the content of the test is decided, the testing method will have to be decided. The type of possible questions are well described in the standard texts and need not be discussed here. The choice will again reflect the course as much as possible. We need to use the testing method that will indicate as near as possible that the student has attained the target ability. For example, it would be inappropriate to ask student who have attained a course which emphasised reading business letters to write a business letter as a test of their successful completion of the course. Similarly, it is not advisable to depend on reading comprehensive on test when checking the success of a course in writing. This is true of other forms of tests, but it needs to be emphasised even more for progress tests for two reasons. *First*, students will tend to concentrate on areas of the course in which they know

they will be tested. *Second*, the correlations between different types of tests that are quoted as a basis for accepting standardised tests, the contents of which do not completely sample the language, may be suspect when *progress tests* are such correlations to be reduced substantially. However, in large institutions or where time is limited, a machine markable test may be used as a common yardstick. If combined with, for example, a teacher's assessment (hopefully, continuous assessment) of areas not covered by the test, and if the students are aware of this, then the disadvantage of using such tests may be avoided while the benefits are retained.

Finally, the questions need to be written and the test booklet designed. These should always be at least double - checked. A second flawn opinion will often see flaws that the first writer cannot see until they are pointed out. The following questions need to be asked:

Are the instructions clear?

Is the content restricted to course content or what might be reasonably thought as a prerequisite to the course?

Are the tasks in openended questions sufficiently defined?
Are they likely to provide the student with an opportunity to show his ability in the relevant skills?

In multiple - choice questions, is there clearly one and only one correct Answer?

Is the marking scheme for the subjectively marked test clear? Does it emphasise areas stressed in the course has stressed communication over accuracy, does the marking reflect?

ANALYSIS

Once the test has been used, there is unfortunately a tendency for it to be put one side. An analysis of the test results can be done fairly rapidly and can provide a lot of information about the students and the course. It can also lead to the development of better progress tests for the future. Often the use of a test can show flaws in it which

cannot be seen by inspection.

One of the first steps in the analysis of the test would be to seek wherever possible the opinions of the teachers and students who have used it. If the *progress test* has been a good one, most students will be satisfied that their results reflect their progress. They will have seen the course reflected in the test and will see that any loss of marks was due to their lack of mastery of sections of the test will usually indicate that either the instructions were not clear, or the question was obscure, or the students did not operate, will also often be able to give helpful advice.

A fairly easy second step is to look at the distribution of marks. As stated in the earlier quotation, the scores on a *progress test* should be clustered towards the top end of the scale if they are not, something, somewhere has gone wrong. There are several possibilities. For example, either the course design needs to be revised or the test needs to be redesigned, or both.

The final step is to investigate each question. It is useful to keep each question on a separate card with a summary showing when it was used and how successful it was. The step may take a little time at first, but will save time later when Making Up New Tests. It will also provide any successor to the present administration a useful guide when they, in turn, have to prepare progress tests.

In the cause of subjectively marked papers, the best check available is to find out how well students did on each question. The scoring scheme will depend a lot on the type of question, and thus hard-and-fast rules are difficult to give. However, generally the question scores should reflect our expectation of the total score, i.e. the students should generally do well. If they have not performed well, a discussion among those involved in teaching the course and preparing the test should help to

identify the problem, and it will either lead to a revision of the questions or a revision of the course.

In the case of multiple-choice tests, an item-analysis can be done. This process is clearly described in several texts. It needs not to take a long time and the information thus obtained can be very useful. However, the item-analysis as described in most texts was developed for checking items in standardised testing and, if used in the same way for progress tests, will not help to develop the type of test that is needed. Thus, the process will be described in detail and any differences of approach needed will be pointed out.

The first step is to rank all the answer booklets according to the total scores obtained by the students. The lower and upper 27% are then set aside for analysis. There should be an equal number of papers in the upper and lower piles. Thus, if there were 100 candidates, there should be 27 papers in each pile. Some times there are several papers with the same score at the point where we have to make the cut - off. For example, there might be in this group of 100, 24 students with scores above 88% and 5 students with scores of 88%. We would then choose 3 of those 5 randomly for the upper group. We could then prepare a table for each item as follows:

	A *	B	C	D	NA	Total
U	24	0	0	3	0	27
L	22	3	0	1	1	27

- | | | |
|------------|---|---------------------------|
| Key A to D | - | Choices in the item. |
| * | - | Indicates correct answer. |
| NA | - | No Answer. |
| U | - | Upper group. |
| L | - | Lower group. |

From these tables we can easily work out two statistics: The difficulty coefficient (= Simply the proportion of people who got the item right) and the discrimination Coefficient (= a Figure which tells us how well the item discriminates between the upper and lower group). The formulas are:

$$\begin{aligned}\text{Difficulty Coefficient} &= \frac{UR+LR}{2n} \\ \text{Discrimination Coefficient} &= \frac{UR-LR}{2n}\end{aligned}$$

Key:

- UR - The number of people in the upper group who got the items right.
- LR - The number of people in the lower group who got the items right.
- n - The number of people in one of the groups.

(N.B. There is a more complicated and accurate formula for discrimination for those contemplating using a computer).

For the above example, difficulty is

$$\frac{24+22}{2 \times 27} = \frac{46}{54} = 0.85 \text{ (to 2 decimal places).}$$

In standardised testing, this would generally be regarded as too easy an attempt is made to spread the scores along the whole scale. Items that all or nearly all candidates get rights or wrong do not help to do this, and so items with difficulties over 0.80 and less than 0.20 are rejected from such tests. However, in progress test we except the items to represent areas the students have mastered, and thus easy items should be retained. On the other hand, items which majority of students get wrong are suspect. Any item that has a difficulty of less than 0.50

may have been badly written or may test on area not covered sufficiently in the course. In the first case it should be rejected. In the second, either the course should be received or the item rejected.

The discrimination coefficient for the above example would be

$$\frac{24 - 22}{27} = \frac{2}{27} = 0.07 \text{ (to decimal places)}$$

Low discrimination would be regarded as unsatisfactory. Low discrimination indicates low agreement with the other items in the test. As much agreement as possible is needed between items to help spread the candidates along the whole scale. Thus, generally items with a discrimination of less than 0.20 are rejected for such tests. However, in *progress testing* this is not such an important consideration. Indeed, it can be proved mathematically that, when using the formula given here, an item with a difficulty over 0.20 cannot have a discrimination above 0.2, and a question that students all get right can only have a discrimination 0.00. Further more, it is unlikely that an item will have the maximum discrimination theoretically possible. Thus, few items with difficulties over 0.80 will have discrimination over 0.20. In a *progress test*, as we wish for to retain these easy items, we will have to accept lower discrimination Coefficients.

However, the discrimination coefficients remain useful. In *progress testing*, as in standardised testing, we use the total score rather than the score and individual questions, and so it is still important that particular items do not work against the rest of the test. If more of the lower group get the answer right than the upper group, then the discrimination coefficients will be negative and the item will be working against the rest of the test. Such an item should be rejected or revised. A suggested

approach is to reject all items with a negative discrimination difficulty coefficient of less than 0.8 and a discrimination coefficient of less than 0.20.

In this way we would build a test that could identify the weaker students while continuing to have a cluster of scores towards the top of the scale.

Following this we should check the distracters in the table previously given. The distracters are the incorrect choices (B, C and D in the example given). Distracter B would be acceptable in any test as more of the lower group cross it, and it thus helps to identify the lower group. Distracter D would be rejected or revised as more of the upper group cross it, and it thus confuses the issue. Distracter C is more difficult. It would be rejected in standardised testing as it does not help to discriminate and is thus just a waste of candidates reading time. However, in *progress testing*, if it represents a common error which students completing the course have mastered, it is worth keeping if on inspection, however, it proves to be a distracters that is so ridiculous that no one would think of choosing it, it is worth revising.

The final stage is to go through all the reject terms and try to establish why they are rejects. Many will simply be badly written, but others will provide keys to where the course is failing to cover certain areas or failing to clear up misunderstandings or where it is decided that the course is at fault, the course can be revised and the item retained.

The above discussion has been an attempt to suggest an approach to the preparation and analysis of *progress tests* that emphasise such tests as *a tool for the educational administrator or teacher, a tool to help check and revise the course so that the aims of the course may be realistic for both students and teachers*. Such tests cannot function properly if they are divorced from those

who are responsible for the courses used wisely, *progress tests* will lead to better designed courses. *Progress tests* are an adjunct to courses. They should never supersede them.

Albert H.Marck Wardt, the author of 'The place of literature in the teaching of English as a second or foreign language' (1978) concludes in his own words (1981): "It is wholly justifiable concerned with the language per se and with taking every possible advantage of the systematic study of language to facilitate the learning process, there is a danger of over-looking or under-evaluating some of the uses to which language may be put, among them is its function as a literacy medium".³

REFERENCES

1. Heaton, J.B. (1975), "Writing English Language Tests", London: Longman, p.239.
2. Allen M.J., "Testing Methodology", Concise Encyclopaedia of Psychology, (ed.) Raymond J.Conssni: Pp.1114-15.
3. Marck Wardt, A.H., (1981), "What Literature to Teach: Principles Selection, and Class Treatment, Forum", Vol.No.XIX.