

COMMON TEST TECHNIQUES

Vasylevskaya V.L.

Белорусский государственный университет

Лыповка V.G.

BA PБ

Test techniques are means of eliciting behaviour from candidates that will tell us about their language abilities. Test techniques:

- will elicit behaviour which is a reliable and valid indicator of the ability in which we are interested;
- will elicit behaviour which can be reliably scored;
- are as economical of time and effort as possible;
- will have a beneficial backwash effect, where this is relevant.

The common test techniques are: multiple choice, Yes/No and True/False, short answer, gap filling items.

Multiple choice items

Multiple choice items take many forms, but their basic structure is as follows.

There is a *stem*:

Enid has been here _____ half an hour.

and a number of *options* - one of which is correct, the others being *distractors*:

A. during B. for C. while D. Since

It is the candidate's task to identify the correct or most appropriate option. Perhaps the most obvious advantage of multiple choice is that scoring can be perfectly reliable. Scoring should also be rapid and economical. A further considerable advantage is that, since in order to respond the candidate has only to make a mark on the paper, it is possible to include more items than would otherwise be possible in a given period of time, this is likely to make for greater test reliability. Finally, it allows the testing of receptive skills without requiring the test taker to produce written or spoken language.

The advantages of the multiple choice technique were so highly regarded at one time that it almost seemed that it was the only way to test. While many laymen have always been sceptical of what could be achieved through multiple choice testing, it is only fairly recently that the technique's limitations have been more generally recognised by professional testers. The difficulties with multiple choice are as follows.

The technique tests only recognition knowledge

If there is a lack of fit between at least some candidates' productive and receptive skills, then performance on a multiple choice test may give a quite inaccurate picture of those candidates' ability. A multiple choice grammar test score, for example, may be a poor indicator of someone's ability to use grammatical structures. The person who can identify the correct response in the item above may not be able to produce the correct form when speaking or writing. This is in part a question of construct validity; whether or not grammatical knowledge of the kind that can be demonstrated in a multiple choice test underlies the productive use of grammar. Even if it does, there is still a gap to be bridged between knowledge and use; if use is what we are interested in, that gap will mean that test scores are at best giving incomplete information.

Guessing may have a considerable but unknowable effect on test scores

The chance of guessing the correct answer in a three-option multiple choice item is one in three or roughly thirty-three per cent. On average we would expect someone to score 33 on a 100-item test purely by guesswork. We would expect some people to score fewer than that by guessing, others to score more. The trouble is that we can never know what part of any particular individual's score has come about through guessing. Attempts are sometimes made to estimate the contribution of guessing by assuming that all incorrect responses are the result of guessing, and by further assuming that the individual has had average luck in guessing.

While other testing methods may also involve guessing, we would normally expect the effect to be much less, since candidates will usually not have a restricted number of responses presented to them (with the information that one of them is correct).

If multiple choice is to be used, every effort should be made to have at least four options (in order to reduce the effect of guessing). It is important that all of the distractors should be chosen by a significant number of test takers who do not have the knowledge or ability being tested. If there are four options but only a very small proportion of candidates choose one of the distractors, the item is effectively only a three-option item.

The technique severely restricts what can be tested

The basic problem here is that multiple choice items require distractors, and distractors are not always available. In a grammar test, it may not be possible to find three or four plausible alternatives to the correct structure. The result is often that the command of what may be an important structure is simply not tested. An example would be the distinction in English between the past tense and the present perfect. For learners at a certain level of ability, in a given linguistic context, there are no other alternatives that are likely to distract.

It is very difficult to write successful items

A further problem with multiple choice is that, even where items are possible, good ones are extremely difficult to write. Professional test writers reckon to have to write many more multiple choice items than they actually need for a test, and it is only after trialling and statistical analysis of performance on the items that they can recognise the ones that are usable. Multiple choice tests that are produced for use within institutions are often shot through with faults. Common amongst these are: more than one correct answer; no correct answer; there are clues in the options as to which is correct (for example the correct option may be different in length to the others); ineffective distractors. The amount of work and expertise needed to prepare good multiple choice tests is so great that, even if one ignored other problems associated with the technique, one would not wish to recommend it for regular achievement testing (where the same test is not used repeatedly) within institutions. Savings in time for administration and scoring will be outweighed by the time spent on successful test preparation

Cheating may be facilitated

The fact that the responses on a multiple choice test (a, b, c, d) are so simple makes them easy to communicate to other candidates non-verbally. Some defence against this is to have at least two versions of the test, the only difference between them being the order in which the options are presented.

All in all, the multiple choice technique is best suited to relatively infrequent testing of large numbers of candidates. This is not to say that there should be no multiple choice items in tests produced regularly within institutions. In setting a reading comprehension test, for example, there may be certain tasks that lend themselves very readily to the multiple choice format, with obvious distractors presenting themselves in the text.

YES/NO and TRUE/FALSE items

Items in which the test taker has merely to choose between YES and NO, or between TRUE and FALSE, are effectively multiple choice items with only two options. The obvious weakness of such items is that the test taker has a 50% chance of choosing the correct response by chance alone. There is no place for items of this kind in a formal test, although they may well have a use in assessment where the accuracy of the results is not critical. True/False items are sometimes modified by requiring test takers to give a reason for their choice. However, this extra requirement is problematic, first because it is adding what is a potentially difficult writing task when writing is not meant to be tested (validity problem), and secondly because the responses are often difficult to score (reliability and validity problem).

Short-answer items

Items in which the test taker has to provide a short answer are common, particularly in listening and reading, tests.

Advantages over multiple choice are that:

- guessing will (or should) contribute less to test scores;
- the technique is not restricted by the need for distractors (though there have to be potential alternative responses);
- cheating is likely to be more difficult;

- though great care must still be taken, items should be easier to write.

Disadvantages are:

- responses may take longer and so reduce the possible number of items;
- the test taker has to produce language in order to respond;
- scoring may be invalid or unreliable, if judgement is required;
- scoring may take longer.

The first two of these disadvantages may not be significant if the required response is really short (and at least the test takers do not have to ponder four options, three of which have been designed to distract them). The next two can be overcome by making the required response unique (i.e. there is only one possible answer) and to be found in the text (or to require very simple language).

Gap filling items

Items in which test takers have to fill a gap with a word are also common. Gap filling items for reading or listening work best if the missing words are to be found in the text or are straightforward, high frequency words which should not present spelling problems.

Gap filling items can also work well in tests of grammar and vocabulary. But it does not work well where the grammatical element to be tested is discontinuous, and so needs more than one gap. An example would be where one wants to see if the test taker can provide the past continuous appropriately.

When the gap filling technique is used, it is essential that test takers are told very clearly and firmly that only one word can be put in each gap. They should also be told whether contractions (I'm, isn't, it's, etc.) count as one word. Counting contractions as one word is advisable, as it allows greater flexibility in item construction.

Gap filling is a valuable technique. It has the advantages of the short answer technique, but the greater control it exercises over the test takers means that it does not call for significant productive skills. There is no reason why the scoring of gap filling should not be highly reliable, provided that it is carried out with a carefully

constructed key on which the scorers can rely completely (and not have to use their individual judgement).

These are common testing techniques. Each of them has its own advantages and disadvantages which should be taken into account developing a test.

REFERENCES

1. Fulcher, G. An English language placement test: issues in reliability and validity. / G.Fulcher. - Cambridge University Press, 2007.- 114-138 p.
2. Hughes, A. Testing for Language Teachers / A.Hughes. - Cambridge University Press, 2003.- 81 p.