

Multiple-choice Item and Its Backwash Effect on Language Teaching in China

Shouyuan Luo

Bio-chemical Engineering Institute of Beijing Union University, Beijing, China

Email: luosy1020@sina.com

Xiaodong Zhang

Bio-chemical Engineering Institute of Beijing Union University, Beijing, China

Abstract—At present, the multiple-choice item is one of the widely used items in language testing, especially in large-scale language testing. In recent years, however, many people criticize this kind of item because of its negative backwash effect on foreign language teaching and learning. This paper seeks to analyze it and its backwash effect from several aspects and angles so as to have a scientific attitude towards it.

Index Terms—multiple-choice item, backwash effect, language teaching

I. GENERAL INTRODUCTION TO MULTIPLE-CHOICE TEST

During the 1960s, influenced by the aural-oral approach and the structuralist linguistic approach, language testing bid farewell to the approaches which were not scientific and introduced psychometric-structuralist linguistic approaches. Lado and his followers (1961) hold that language consists of different parts such as phonetics, vocabulary and syntax, and different skills such as speaking, listening, reading and writing. These different parts and skills can be broken down into their component parts and those parts tested separately and adequately (Farhady, 1979). Such a rationale is derived from structural linguistics and behavioristic psychology which claim that behavior is in essence the sum of its component parts. It is claimed that by adequately sampling the component parts of language, validity can be achieved. It is just under such background that multiple-choice item came into being.

Very often, the multiple-choice item is made up of two parts. The initial part of each multiple-choice item is known as the stem; the choices from which the students select their answers are referred to as options. One option is the correct answer, while the other options are distractors.

Multiple-choice item is now very popular and widely used in national and international examinations such as NMET, CET4, CET6, EPT in China, and TOEFL, GRE, GMAT, BEC, IELTS in the United States and in England. However, the voice of criticism against using this kind of item is increasingly growing, for more and more people are worried about the negative backwash effect caused by multiple-choice item. In order to make an objective assessment of the item, it is necessary for us to discuss such a kind of item and its backwash effect on teaching in this paper.

II. ADVANTAGES AND DISADVANTAGES OF MULTIPLE-CHOICE TEST

In analyzing the advantages and disadvantages of multiple-choice item, we should take into consideration the principles of language testing. Generally speaking, the advantages and disadvantages are mainly as follows:

A. Advantages of Multiple-choice Item

There are a series of advantages with multiple-choice item. Firstly, it is easy for scorers to go over examination papers. Since it usually has only one correct answer, or at least, a limited number of correct answers, so it can be scored mechanically, and therefore, the reliability of scorers can be ensured.

Secondly, besides its high reliability of scoring, multiple-choice item has the strengths of saving time, effort and money in carrying out large-scale examinations. Meanwhile, within the same amount of time, the exam which adopts this kind of item can embody more items in a test paper. In doing so, the reliability of the exam can be greatly improved. Moreover, it is quite convenient for statistic analysis and can also provide more feedback for language teaching and learning, which makes this kind of test more scientific.

Thirdly, the range of the multiple-choice item is wide enough. It can be employed in any disciplines, which strives for the process of dealing with the problems, such as mathematics, chemistry and physics. In English teaching and learning, whether practical courses or theoretical courses such as linguistics, literature, this kind of item is widely used in tests. Viewed from linguistic level, the multiple-choice item may be employed at the word level, sentence level and discourse level. And viewed from the level of ability, although it is concerned mainly with the ability of memory and comprehension, at times it can be beyond this scope.

Fourthly, it has a good diagnostic function. In multiple-choice items, the content which is designed for testing can be

divided into several items to be tested. Therefore, teachers can find out problems in time by means of analyzing mistakes made by the testees in each item and then put forward approaches to give remedial teaching to the students.

Fifthly, multiple-choice items can provide a useful means of teaching and testing in various learning situations (particularly at the lower levels) provided that it is always recognized that such items test knowledge of grammar, vocabulary, etc, rather than the ability to use language.

B. Disadvantages of Multiple-choice Item

There are many disadvantages with multiple-choice item, firstly, frequently it does not lend itself to the testing of language as communication. The process involved in the actual selection of one out of four or five options bears little relation to the way language is used in most real-life situations. Appropriate responses to various stimuli in everyday situations are produced rather than chosen from several options.

Secondly, the objective test of the multiple-choice type does encourage guessing. It is quite obvious that the factor of guessing is one that can't be excluded in the process of choosing the key in a test paper. At present, most of the multiple-choice item has four alternatives, therefore, the chance of a student guessing an answer is 25%, even if candidates are ignorant of the four options, and also the chance will surely be higher if they can eliminate one or two of the distractors.

Thirdly, the process of constructing items is quite complex. It is common knowledge that the process of constructing subjective items is easier than that of constructing multiple-choice item. For example, the type of writing item can generally measure candidates' ability to write if the title of writing is suitable and personal and subjective factors are controlled appropriately. On the contrary, the quality of multiple-choice item is hard to ensure because of this complex process of item writing. At first sight, the construction of multiple-choice item is quite simple, but in fact, it is rather difficult to control its item and alternatives. A number of questions should be taken into consideration, for example, which linguistic elements should be tested, how to make pretest, what kind of data analysis must be made after pretest, how to design item and key in each item, how to design plausible distractors, and so on, all of which really cost much time and effort.

Fourthly, it is easier for candidates to cheat in multiple-choice item. The reason for this phenomenon is that candidates are only required to choose the key from four alternatives, so it is quite convenient for some candidates to copy other candidates' answers. Even though from the test papers the scorers can make sure that some candidates have cheated in the exams, however, it is not easy to collect strong evidence in order to punish them.

III. BACKWASH EFFECT ON LANGUAGE TEACHING

From what we have discussed above, we can see clearly that the widely application of multiple-choice item in large-scale examination produces negative backwash effect on language teaching and learning. For example, many candidates can obtain high marks in a test containing multiple-choice item, but fail to communicate with other people in English freely. Likewise, a number of candidates who have passed CET4 or CET6 are eliminated in English interview when they apply for jobs in foreign companies. Clearly, in order to get high marks, both teachers and students are very interested in looking for test-taking strategy rather than the training of language itself. As we know, the aim of exam is to test teaching quality, promote teaching quality and to deepen the reform of foreign language teaching and learning. If candidates use the wrong means to pass the exam, then such kind of testing will be meaningless.

Since it is the product of the second stage in the development of language testing, multiple-choice item focused on testing receptive skills of language rather than productive skills of language. Just as testing expert Alderson put, "there is evidence that students taking multiple-choice tests can learn strategies for taking such tests that 'artificially' inflate their scores." (Alderson et al, 2000, p.45) Obviously, such kind of item gives rise to negative backwash effect on language teaching and learning. Therefore, to prepare for this kind of item is not the best way to improve candidates' language ability.

IV. APPLICATION OF MULTIPLE-CHOICE TESTS TO TEACHING TO REDUCE THE NEGATIVE BACKWASH EFFECT

In the improvement of the quality of a test, such as the multiple-choice test. The most important thing is to increase the reliability and validity. Reliability refers to the fact that if the same test is given to the same subject or matched subjects on two different occasions, it yields similar results whereas validity of a test is the extent to which it measures what it is supposed to measure and nothing else. In other words, it means that language testing can accurately reflect candidates' ability to use the language. Viewed from the level of development of language testing, the reliability and validity of a test often contradict each other. The testing methods with high reliability such as multiple-choice item are highly reliable but their validity may be low. The ideal condition is hard to achieve, although the validity can be improved by means of designing the items carefully. On the contrary, testing methods with high validity, such as composition, oral test, are highly valid, but usually the reliability is not high. And because of this factor, it is not easy to guarantee objectivity and consistency of scoring. Some measures may be employed to increase the validity of this kind of test, however, similarly, the ideal extent is quite difficult to attain. Clearly, this is the problem facing foreign language testing. So for modern language testing, large-scale language testing in particular, we can only try to find the best

balance between reliability and validity. To reduce the negative backwash effect in a test containing multiple-choice item, the test paper should contain two essential parts: objective items (mainly multiple-choice item) and subjective items. The combination of objective items and subjective items can make up each other's deficiencies so as to ensure the practicality of operation in large-scale language testing and its communication, and at the same time to make the language testing scientific to the greatest possible.

V. CONCLUSION

As one of the most popular items in language testing, the multiple-choice item deserves to be explored by us not only from theory but also from practice. The fact we should bear in mind is that the multiple-choice item has more disadvantages over advantages. Its big advantage is the objectivity of scoring and economy in terms of manpower, time and material resources. But the complex process of constructing items, especially the writing of distractors, is much more difficult than we can expect. For large-scale tests, however, such as CET and TEM since there are so many candidates taking part in them each year, we have to continue to use it before better and more appropriate items come into being. We have to use it not because it is the best one, but because we have no other choices at present. Therefore, in order to improve the validity of multiple-choice item, the test papers in large-scale test should be made up of two parts, a lot of multiple-choice items on one hand and a few of subjective items (for example, composition item, translation item, paraphrase item) on the other, with the aim of making full use of the strengths of both subjective and objective items to guarantee the quality of test papers. Thus a set of items will have both good reliability and validity, and also produces good backwash effect on teaching.

REFERENCES

- [1] Alderson, J.C. et al. (2000). *Language Test Construction and Evaluation*. Beijing: Foreign Language Teaching and Research Press.
- [2] Backmen, L. et al. (1999). *Language Testing in Practice*. Shanghai: Shanghai Foreign Language Education Press.
- [3] Farhady, H. (1979). The Disjunctive Fallacy between Discrete-Point and Integrative Tests. *TESSOL Quarterly* 13, 40-43.
- [4] Henning, G. (2001). *A Guide to Language Testing: Development, Evaluation and Research*. Beijing: Foreign Language Teaching and Research Press.
- [5] Heaton, T. B. (2000). *Writing English Language Tests*. Beijing: Foreign Language Teaching and Research Press.

Shouyuan Luo was born in Hunan, China in 1965. He received his M. A degree in linguistics from School of Foreign Languages of Southwest Jiaotong University, China in 2005.

He is currently a lecturer in Bio-Chemical Engineering College of Beijing Union University, Beijing, China. His research interests include applied linguistics, British and American literature.

Xiaodong Zhang is currently an Associate Professor of Bio-Chemical Engineering College of Beijing Union University. Her main area of research is applied linguistics and language teaching.