

ADVANTAGES AND DISADVANTAGES OF TESTING FOREIGN LANGUAGE ABILITY

E.V. Riabtseva, A.A. Gvozdeva, L.P. Tsilenko

Department Foreign Languages, TSTU

*Represented by Professor M.N. Makeyeva
and a Member of Editorial Board Professor V.I. Kononov*

Key words and phrases: cloze tasks; cloze elide; c-tests; information transfer technique; multiple – choice questions (MCQs); selective deletion gap filling; short answer questions.

Abstract: In this article the advantages and disadvantages of constructing tests and test items are pointed out to control student's foreign language ability. Here are introduced the most frequently used test items and the principles of their writing.

Nowadays many teachers of foreign languages “make efforts” to use testing to control the students' knowledge of a foreign language. We say “make efforts” because very often the teachers have know a good idea of what any test should consist, what it should measure and how the knowledge of a foreign language should be evaluated.

Communicative language testing has recently introduced another dimension named profile reporting. In order to obtain a full profile of a student's ability in the target language, it is necessary to assess his or her performance separately for each of the different areas of communication: e.g. listening comprehension, speaking and listening, reading, reading and writing (summarizing, etc.) and writing. Furthermore, performance is assessed according to the purpose for which the language is to be used: e.g. academic, occupational, social. The object of the sub-tests (tasks of a test) through which performance is assessed is to indicate the extent of the learner's mastery of the various language skills which he or she will require for a particular purpose. A score or grade is given for each of the skills or areas selected, and an average mark is eventually obtained. This latter mark, however, is only given alongside the various scores which have contributed to it.

Thus, we believe, profile reporting is very valuable for placement purposes, and indeed it is an essential feature of one of the most widely used proficiency tests set in Britain and administered in many countries throughout the world. A student's performance on the various parts of the test can be shown in the form of a simple table or chart, in which the target score appears beside the student's score. To our mind, it makes the process easier to compare a student's performance levels in each area with the required levels.

Besides a test must be practicable: in other words, it must be fairly straightforward to administer. We should remember that it's only too easy to become so absorbed in the actual construction of test items that the most obvious practical considerations concerning the test are overlooked. The time period devoted to the administration of the test is often misjudged even by experienced test writers, especially if the test consists of a number of sub-tests. In such cases we are short of time to

administrate the test, to collect the answer sheets, to read the test instructions, etc. In the case of all large-scale tests, the time to be allowed should be decided on as a result of a pilot administration of the test (i.e. a tryout of the test to a small but representative group of testees).

As to the test instructions to the candidate, it is essential that all instructions are clearly written and that examples are given, since most students taking any tests are working under certain mental pressure. Sorry to say, many teachers prefer to use grammatical terminology that should undoubtedly be avoided, and such rubrics as the following rewritten:

- put the correct pronoun in the blanks;
- choose on of the following verbs to go in each blank space and put it in the correct tense.

Students may be able to perform all the necessary tasks without referring to grammar rules. Thus, the first of the rubrics above should be rewritten and the phrase “words like the following” (followed by examples) be used to replace “pronouns”. The second rubric should have words instead of verbs, and examples should be given so that the students are shown the tense changes they are required to make. This principle does not apply only to grammatical terms: if students are instructed to put a tick opposite the correct answer, the example of what is meant by the word “tick” should be given – e.g. (*I*) the same applies to crosses, circles, underlining, etc.

Now we would like to describe different types of test tasks so that it will help the teachers to construct tests of good quality. The most frequently used types of test tasks are as follows:

- multiple – choice questions (MCQs);
- short answer questions;
- cloze tasks;
- selective deletion gap filling;
- c-tests;
- cloze elide.

A **multiple – choice test item** is usually set out in such a way that the candidate is required the answer from a number of given options, only one of which is correct. The marking process is totally objective because the marker is not allowed to exercise judgement when marking the candidate’s answer; agreement has already been reached as to the correct answer for each item. Selecting and setting items are, however, subjective processes and the decision about which is the correct answer is the matter of subjective judgement on the part of the item writer.

We think that we should point out both the advantages and the disadvantages of MCQs. In MCtests there is almost complete marker reliability candidate’s marks, unlike those in subjective formats, can not be influenced by the personal judgement. The marking, as well as being reliable, is simple and more rapid.

Because items can be pre-tested fairly easily, it is usually possible to estimate in advance the difficulty level of each item and that of the test as a whole. Pre-testing also provides information about the extent to which each item contributes positively towards what the test as a whole is measuring. Ambiguities in wording of items may also be revealed by analysis of the pre-test data and can then be clarified or removed in the test proper. Besides, the format of the MCtest item is such that the intentions of the test writer are clear unequivocal; the candidates know what is required of them. For example, in open-ended formats ambiguities in the wording of questions may sometimes lead to the candidates submitting answers to questions different from those which the examiner had intended to ask.

Another advantage of MCtests is that they avoid the difficulty of performing the writing skill for the candidates. In open-ended tests this may affect accurate

measurement of the trait being assessed has not been established. At the same time some disadvantages of MCtests can be found. There are a number of problems associated with the use of this format. If a candidate gets a MCitem wrong because of some flow in the question, the answer sheet on which he or she writes down his answer will not reveal this fact. In addition we wouldn't know whether a candidate's failure was due to lack of comprehension of the text or lack of comprehension of the question. A candidate might get an item right by eliminating wrong answers, a different skill from being able to choose the right answer in the first place.

The score gained in MCtests, as in true-false tests, may be suspect because the candidate could guess all or some of the answer.

MCtests take much longer time and are more difficult to prepare than more open-ended exams, e.g. compositions. A large number of items have been written carefully by item writer who have been specially trained and these then have to be pre-tested before use in a formal examination. Each item has to be rigorously edited to ensure that:

- there is no superfluous information in the stem;
- the spelling, grammar and punctuation are correct;
- the language is concise and at an appropriate level for candidates;
- enough information has been given to answer the question;
- there is only one unequivocally correct answer;
- the distractors are wrong but plausible and discriminate at the right level;
- the responses are homogeneous, of equal length and mutually exclusive and the item is appropriate for the test.

More than that, there is a considerable doubt about their validity as measures of language ability. Answering MCitems is an unreal task, as in real life one is rarely given four alternatives from which to make a choice to signal understanding. There are questions which require the candidates to write down specific answers in spaces provided on the question paper – **short answer questions**. The technique is extremely useful for testing both reading and listening comprehension. In these test items answers are not provided for the student as in MCQs. Therefore if a student gets the answer right, one is more certain that this has not occurred for reasons other than comprehension of the text. Then, with careful formulation of the questions a candidate's response can be brief and thus a large number of questions may be set in this format, enabling a wide coverage. If the number of acceptable answers to a question is limited it is possible to give fairly precise instructions to the examiners who mark them. For example, activities such as inference, recognition of a sequence, comparison and establishing the main idea of a text, require the relating of sentences in a text with other items which may be some distance away in the text. This can be done effectively through short answer questions where the answer has to be sought rather than being one of those provided. Then, a strong case can be made in appropriate contexts for the use of long texts with short answer formats on the grounds that these are more representative of required reading in the target situation, at least in terms of length. They can also provide more reliable data about a candidate's reading ability.

The main disadvantage to this technique is that it involves the candidate in writing and there is some concern, largely anecdotal, that this interferes with the measurement of the intended construct. Care is needed in the setting of items to limit the range of possible acceptable responses and the extent of writing required. In those cases where there is more debate over the acceptability of an answer e.g., in questions requiring inferencing skills, there is a possibility that the variability of answers might lead to marker unreliability. However, careful moderation and standardisation of examiners should help to reduce this.

In the **cloze** procedure words are deleted from a text after allowing a few sentences of introduction. The deletion rate is mechanically set, usually between every fifth and

eleventh word. Candidates have to fill each gap by supplying the word they think has been deleted. The reader comprehends the mutilated sentences as a whole and completes the pattern. A cloze test given under timed conditions provides valid and reliable indices of student's proficiency if two conditions are met: first, that the used textual material is of the appropriate level of difficulty for the candidates and second, that it contains a sufficient number of deleted items.

It looks as though cloze tests are easy to construct and easily scored if the exact word scoring procedure is adopted, and they are claimed to be valid indicators of overall language proficiency. But despite the arguments introduced in favour of cloze procedure, a number of doubts have been expressed, largely concerning its validity as a testing device. It has often been irritating and unacceptable to students and doubt has been thrown on the underlying assumption that it randomly samples the elements in a text. Taking into account the candidate's results of a cloze test still it has been under a question if this cloze test really tells us about a candidate's language ability.

In the light of recent negative findings on mechanical deletion cloze, increasing support has developed for the view that the test constructor should use a "rational cloze", selecting items for deletion based upon what is known about language, about difficulty in text and about the way language works in a particular text. Linguistic reasoning is used to decide on deletions and so it is easier to state what each test is intended to measure. This technique is better referred to as **selective deletion gap filling** as it is not 'cloze' in the proper sense. To our mind, selective deletion enables the test constructor to determine where deletions are to be made and focus on those items which have been selected a priori as being important to a particular target audience. It is also easy for the test writer to make any alterations shown to be necessary after item analysis and to maintain the required number of items. This might involve eliminating items that have not performed satisfactorily in terms of discrimination and facility value.

But whereas short answer and MCQs allow the sampling of the range of enabling skills, gap filling is much more restrictive where only single words are deleted. Gap filling only normally allows the testing of sentence bound reading skills. If the purpose of a test is to sample the range of enabling skills including the more extensive skills such as skimming, then an additional format to gap filling is essential.

Recently an alternative to cloze and selective deletion gap filling has emerged for testing comprehension of the more specifically linguistic elements in a text. An adaptation of the cloze technique called the **C-test** has been developed in Germany by Klein-Braley (1981-1985) based on the same theoretical rationale as cloze, testing ability to cope with reduced redundancy and predict from context.

In the C-test every second word in a text is partially deleted. In an attempt to ensure solutions students are given the first half of the deleted word. The examinee completes the word on the test paper and an exact word scoring procedure is adopted.

With C-test a variety of texts are recommended, and given the large number of items that can be generated on small texts this further enhances the representative nature of the language being sampled. Normally a minimum of 100 deletions are made and these are more representative of the passage as a whole than is possible under the cloze technique. Also the task can be objectively scored because it is rare for there to be more than one possible answer for any one gap.

At the same time, given the relatively recent appearance of the technique in this form there is little empirical evidence of its value. Most concern has been expressed concerning its public acceptability as a measure of language proficiency. But this technique suffers from the fact that it is irritating for students to have to process heavily mutilated texts and the face validity of the procedure is low.

The next technique, we are going to mention here is **cloze elide** which is generating interest recently is where words which do not belong are inserted into a

reading passage. The candidates have to indicate where these insertions have been made. There is in fact nothing new about this technique. In its earlier form it was known as the intrusive word technique.

In comparison with multiple-choice format or short answer questions the candidate does not have the problem of understanding the question. It has approximately the same item yield as a cloze test. But there is a problem connected with scoring. This process is highly problematic as candidates may, for example, delete items which are correct, but redundant.

In testing both reading and listening comprehension we have referred to the problem of the measurement being “muddied” by having to employ writing to record answers. In an attempt to avoid this contamination of scores several Examination Boards in Britain have included tasks where the information transmitted verbally is transferred to a non-verbal form, e.g., by labelling a diagram, completing a chart or numbering a sequence of events. Here **information transfer techniques** are particularly suitable for testing an understanding of process, classification or narrative sequence and are useful for testing a variety of other text types. It avoids possible contamination from students having to write answers out in full. It is a realistic task for various situations and its interest and authenticity gives it a high face validity in these contexts. But a good deal of care needs to be taken that the non-verbal task the students have to complete does not itself complicate the process. In some tasks students may be able to understand the text but not what is expected of them in the transfer phase. Besides, there is also a danger of cultural and educational bias. Students in certain subject areas may be disadvantaged, e.g., some students in the social sciences may not be as adept in working in a non-verbal medium as their counterparts in science disciplines.

We believe that a teacher is absolutely free in his choice of test items or tasks. It’s up to him to decide which of them are acceptable for checking target linguistic abilities of his students. We wanted to indicate the advantages and disadvantages of some test techniques to help the test-constructors to improve their test in future.

Literature

1. Heaton J.B. Classroom Testing / Longman Keys to Language Teaching: Longman, 1990. Pp. 96–104.
2. Weir Cyril J. Communicative Language Testing. Prentice Hall, 1990. Pp. 37–51.
3. Heaton J.B. Writing English Language Tests. New ed. London & New York: Longman, 1988. Pp. 159–173.

Преимущества и недостатки тестового контроля владения иностранным языком

Е.В. Рябцева, А.А. Гвоздева, Л.П. Циленко

Кафедра иностранных языков, ТГТУ

Ключевые слова и фразы: выборочные клоуз–задания; задания множественного выбора; задания, требующие краткого ответа; избыточные клоуз–задания; клоуз–задания; «с» тестовые задания; трансформация информации.

Аннотация: В статье указывается на преимущества и недостатки построения тестов и различных типов тестовых заданий. Представлены наиболее часто используемые тестовые задания и описаны принципы их создания.

Vorteile und Nachteile der Testenkontrolle vom Fremdsprachebeherrschen

Zusammenfassung: Es werden die Vorteile und die Nachteile der Testenstruktur und verschiedener Arten der Testenaufgaben gezeigt. Im Artikel sind die am öftesten verwendenden Testenaufgaben vorgestellt und es sind die Prinzipien ihrer Bildung beschrieben.

Avantages et désavantages du contrôle de l'apprentissage des langues étrangères à l'aide des tests

Résumé: Sont indiqués les avantages et les désavantages des tests et des devoirs. Dans l'article sont présentés les devoirs les plus utilisés dans les tests et sont décrits les principes de leur création.
