С.В. Струценко

Types of tests and testing

Let's consider proficiency tests, achievement tests, diagnostic tests and placement tests. Proficiency tests measure learners' language ability regardless of the training they may have had or the vocabulary and topics they may have studied. Proficiency tests are not based on the contents of a language course but rather on the general knowledge of the target language and culture. Achievement tests are directly related to the language courses taught to the examinees. The purpose of achievement tests is to judge upon the success of individual learners or groups in achieving the objectives of the language course. Achievement tests are always «course related» and refer to course contents and objectives. Diagnostic tests identify students' strengths and weaknesses. They provide the teachers with the information on what further teaching is necessary and what problems the students might have in coping with the instruction demands. Placement tests provide information that helps to place the students at the most suitable stage of the teaching curriculum, bearing in mind their level of language achievement so far [1].

Multiple-choice tests have a «stem» (the basic and unaltered part of the sentence) and a number of «options», only one of which is correct. The other options are wrong in the particular context and are called «distracters». The advantages of multiple-choice tests are reliable and economical scoring. A test can include a fairly long list of items and increase the reliability of procedure, thus decreasing the randomness of the results. The disadvantages of the multiple-choice test are that it checks only recognition knowledge. Guessing can have an effect on the scores. Plausible distracters (i.e., those which look correct) are not always available and this makes test writers include «fool-proof «distracters. The «correct answer» can in quite a few cases be questioned.

Gap filling refers to tasks where the test-takers are given separate sentences with some words or phrases deleted. The task is to restore the missing words. In these tests answer keys can sometimes have more than one answer for a space. Some missing words can have a structural value for the sentence (e.g. prepositions or conjunctions). Other words can have full lexical meaning. Sometime a list of the words to be used in filling in the gaps is given to the test-takers. In such cases the number of words in the list is usually larger than the number of gaps in the text [2].

Cloze procedure involves deleting a number of words in the whole text, requiring the test-takers to restore the original words. The procedure is based on the assumption of gestalt psychology that human psyche tends to «complete the closures» in the elements of the surrounding world that is being perceived. This makes the cloze procedure similar to cognitive processes. The first short passage of the text is usually left not mutilated. After this brief «lead-in» approximately every seventh word is deleted. Disadvantages of the test are that some words prove more difficult to restore than others. There could be more than one answer for any one

gap. It is not always easy to say what language area (grammar, lexis, prepositions, etc.) or skills (receptive or productive) are actually measured by the cloze procedure.

Matching is a test format when students are given a list of items, which they have to match with another list of words, phrases, sentences, paragraphs or visual images. The disadvantage of it is that once the test-takers have successfully made a sufficient number of matches, the remaining pairs can be guessed more easily and the last pair will be correct by default.

Information transfer means that the test-takers have to transfer material from the text to a chart, table or map. Such tasks are sometimes performed in real life settings (arranging information during a lecture or finding out the details of the train timetable) and therefore can be viewed as authentic.

Ordering tasks (sequencing) are typically used to test the ability of arranging a discourse (stretch of an oral or written utterance) in a logically organized chain. The students are presented with a mutilated text, in which the order of elements is jumbled. The task is to restore the logical sequence. The problem with «sequencing tasks» is that in some cases there can be more than one way of ordering the textual elements.

Editing tests consist of sentences or text passages, in which errors have been introduced. Test-takers are to identify the errors and to write out the missing or erroneous elements at the end of each line. The main difficulty with this type of test is to ensure that an element deemed as erroneous by test writers should indeed be erroneous with all exceptions excluded [3].

Remodeling tests mean that the learners have to rewrite original sentences starting with the given words or using a given word so that the meaning of the original sentence does not change.

Test analysis is necessary in order to process the results and to draw conclusions about the learners' performance. There are two chief methods of marking the test results. They are the error-count method and the analytic method. The error-count method involves counting up the number of errors and deducting this number from the maximum total scoring for the test. The analytic method of marking is useful because it informs the test-takers about their performance against certain criteria. Possible criteria for assessing a piece of free writing is given below:

5 points. Clear message. Coherent discourse (very logical writing). Topical vocabulary. Perfect cohesion (excellent use of connectives). Correct grammar.

4 points. Clear message with an occasional deviation from the subject. On the whole coherent discourse with some parts «falling out». Mostly topical vocabulary. Use of connectives. Adequate grammar.

3 points. A message is recognizable. Some passages do not belong to the subject. Some topical vocabulary is used. Some connectives are erroneous. There are grammar errors.

2 points. The message is poorly conveyed. Only a small part of the writing actually carries the subject. Much topical vocabulary is missing. Connectives are mostly out of place. Grammar errors make the comprehension difficult.

1 point. The writer has missed the message. The whole text is off the given subject. There is no use of the topical vocabulary. Textual grammar makes the piece incomprehensible [2: 108-117].

An important consideration in test analysis is discrimination, i.e. the ability of the test to signal the individual differences in the learners' language performance. Discrimination can be measured with the help of the formula:

$$\text{Discrimination} = \frac{\text{Upper half of the group} - \text{Lower half of the group}}{\text{Total number of students}}$$

In this formula, numerator shows the number of students in the half of the group with the lower scoring, and the number of students in the half of the group with the higher scoring. Denominator shows the total number of students in the group. If D (discrimination) is 0, the discrimination capacity of the test is very high and it definitely shows the difference in the performance of the test-takers. It is neither too difficult, nor too easy. If D is approaching + 1, there is little or no discrimination in individual language performance and the test produces very high results, thus being too easy. This test is useless in assessing the learners' performance. If D is approaching – 1, then the test is definitely too difficult and can't be used as a measuring tool [4].

Testing writing pursues the goal of testing the ability to write. The following tasks are taken from popular tests: «You are asked to give a talk about your life abroad. Write down what you will speak about», «Write your own thoughts on the saying 'Brevity is the mother of talent'«, «Being born into a wealthy family: a curse or a blessing?», etc. Testing writing can include the operations of completing, describing, explaining, comparing and contrasting, reasoning on paper in written form. Completion is done in the forms that may turn out to be useful in the real world (e.g. the application form).

Describing format is used to test students' abilities to describe their message to an addressee. E.g. You have got a picture that you want your friend to use as a logo. There is no way you can show the picture. Write a description of the image that you want to be used as a logo. Explaining format is often used in the «letter-of-complaint» task. The subject of testing in this case can be the ability of the test-takers to describe the matter of complaint and the ways to put it right. Contextual format means that the examinees are given brief notes of a public address and the task is to «unfold» these brief entries into full text. Written response format consists of the task to read a letter from an official or a clerk and to write a response, focusing on the chunks of the text, which have been underlined. The «legend» for the response is provided. This can be «biography data», «previous letter sent to the clerk by the test-taker», «background situation for this exchange of letters», etc. Comparing and contrasting format is used to test the ability to write a «balanced argument», in which all «pros» and «cons» are equally represented. Reasoning is tested in writing essays.

Testing oral ability has the objective of measuring language performance in oral interaction. Oral tests represent a set of oral tasks that the candidates are to be able to perform. These tasks are expected to elicit behavior, that reveals the oral ability of the test-takers and that can be scored validly and reliably. Oral behavior

that is to be scored includes expression (thanks, apologies, attitudes, opinions), manipulation (directions, instructions, persuading, advising, warning, complaining, banning and permitting), eliciting (information, clarification, explanation), narrating (sequence of events, story-telling, eye-witness accounts). Oral test types can be a monologue, a dialogue and a multi-participant interaction. Criteria for scoring the performance at a certain level of language acquisition include accuracy, appropriacy, range (of the language), flexibility, size [5].

There are three general formats for testing oral ability: interview, interaction with peers and response to an image, written or heard text. Interviews can be structured and unstructured. Structured interview means

having predetermined questions with limited possible answers. Unstructured interview means that the questioning is guided by the responses to the previous questions. Interviews can be administered on a certain topic or cover a wider range of themes. Interaction can be arranged between two test-takers discussing a topic, giving explanations or making plans. E.g. «Discuss how you make an omelet», «Describe a stapler», «Explain how to use a public pay-phone.» Candidates can be asked to assume roles in a particular situation. A protracted exchange between the two candidates can be stimulated by the following task: «You want your mother to increase your pocket money, but she is reluctant. Try to мaкe your mother change her mind.» Discussion between the candidates can also be an important source of information about their language performance. E.g. «Your school can spend money on one of the following items to improve facilities: video equipment, a minibus, computer equipment, a sauna, library resources. Discuss with your partner all the advantages and disadvantages of each suggestion and try to reach an agreement on the most suitable decision». Response to a picture can take the form of description and comment (a picture is expected to be either vague in its form and reference, or deep in its message). A response can be given to a text for reading.

In order to assess the fluency of oral ability, the following scales can be used: background knowledge, vocabulary sufficiency, grammar adequacy, message communication, interactive skills, logic of the utterance, fluency. Grades are given on every scale. Only when examines can be relied upon to score video-recordings with consistent accuracy, should they be entrusted with «live scoring» [6].

Testing reading. The following operations are necessary to test if the task is to evaluate candidates' language performance in reading: scanning the text to locate specific information, skimming the text to obtain the gist, identifying examples in support of an argument, restoring the sequencing relations between parts of the text, inferring from the text. Techniques for testing reading include multiple choice, sequencing, cloze, information transfer, comments, drawing conclusions. In the multiple choice tests the candidates give evidence of successful reading by ticking off one out of a number of alternatives (usually out of three, four or five] alternatives). True-False questions are just a variety of the multiple choice format. E.g. «Choose the picture (A, B, C, D), which the following text describes.» In the cloze test the task is to complete the gaps in the text.

Information transfer demands on the candidates to supply information from the text in a table, map, picture, etc.

Restoring the sequencing relations between the parts of the text can be done in at least two ways. Candidates can be given a rambled order of text fragments, which they are expected to rearrange in the correct sequence. As an alternative, the candidates are given a text and a separate passage. Their task is to decide where this separate passage belongs in the text.

Comments usually require knowledge of the world, The candidates are given a text and asked to explain the meaning of certain words or circumstances. They have to display their historical, social or cultural back- ground.

Testing listening. All the tasks that have been set tor listening should be. done «while-listening». The tape is usually played twice. A couple of minutes are allowed after the tape has been stopped for the test-takers to go over their answers once again. The candidates are to perform the following operations: listen for specific details, obtaining a gist of what they hear, follow directions, follow instructions, interpret the text. Techniques for testing listening can include multiple choice, sequencing, information transfer, following instructions and directions, comments, drawing conclusions.

Testing grammar can be done with multiple choice items, paraphrase (say it differently using a different beginning of the sentence or a different word), sentence completion, cloze procedure (e.g. testing the knowledge of prepositions or articles), error correction, etc.

Testing vocabulary is essential to demonstrate linguistic skills. Test tasks can include synonyms. (Choose the alternative ABCD, which is closest in meaning to the given words. E.g. Gleam: a) shine, b) glitter, c) glare, d) glittering. Answer key: «gleam» has the implication of feeling joy. That is why «shining» is closest in meaning to it.) Definition task can have a multiple choice format («Loathe» means: a) to dislike intensely, b) to hate the look of, c) to feel sick at the smell of. Answer key: «to dislike intensely» is the best definition). Gap filling or cloze procedure can also be used in testing vocabulary. Tasks with pictures can be as follows: «Write down the names of the objects that you see in the picture.» This technique is restricted to concrete nouns. Another technique is to elicit words from definitions (E.g. ... is a person who performs operations on patients. ... is what becomes of boiling water. ... is what a coward needs. Answer key: surgeon, steam, courage) [3: 16-17].

Testing can be applied in attempts to measure the language aptitude of learners. One of the techniques to measure the language analytic and encoding ability is to give the learners examples in an artificial or rare language.

Библиографический список

1. *Bachman, L.* Fundamental Considerations in Language Testing [Text] / L. Bachman. – OUP, 1990.
2. *Heaton, J.* Writing English Language Tests [Text] / J. Heaton. – Longman, 1988.
3. *Alderson, Ch.* Language Test Construction and Evaluation [Text] / Ch. Alderson, C.Clapham, D.Wall. – CUP, 1996.
4. *Bachman, L.* Language Testing in Practice [Text] / L. Bachman, A. Palmer. –

OUP, 1997.
5. *Huges, A.* Testing for Language Teachers [Text] / A. Huges. – CUP, 1996.
6. *Heaton, J.* Classroom Testing [Text] / J. Heaton. – Longman, 1999.