



УДК 811.112.2

CRITERIA FOR EVALUATING QUALITY OF FOREIGN LANGUAGE TESTS**КРИТЕРІЇ ОЦІНЮВАННЯ ЯКОСТІ ТЕСТІВ З ІНОЗЕМНОЇ МОВИ****Yaremenko I.A. / Яременко І.А.***s.ph.s., as.prof. / к.філос.н., доц.*

ORCID: 0000-0001-7144-6956

Tokar L.O. / Токар Л.О.

ORCID: 0000-0002-9448-5883

*Dnipro University of technology, Dnipro, Pr. Dmytra Yavornytskyogo, 19, 49005**Національний технічний університет «Дніпровська політехніка»,**Дніпро, пр. Дмитра Яворницького, 19, 49005*

Abstract: *Different models of linguistic competences are analysed. Certain principles of the development of the 1960s-1970s language tests are considered with the emphasis on their key problems and disadvantages. Test models, based on the principles of communicative competence, their types, purposes, and application spheres are studied. Special attention is paid to the issues concerning development of the criteria for quality assessment of foreign language tests, i.e. validity, reliability, objectivity and standardization. Conclusions are made as for the importance of test use while organizing both in-class and independent student's work.*

Keywords: *language tests, models of communicative competence, criteria for proper assessment of language test results, criteria for evaluating quality of foreign language tests, content validity, reliability, objectivity*

Introduction.

Testing as the knowledge quality control is the integral part of the system of foreign language learning. Contrary to the process of foreign language studying, which purpose is to form foreign-language speech habits and skills, test control is aimed first of all at assessment of their level of development. Thus, regulation of that process is the main task of the control of knowledge level and quality while foreign-language learning. Being the integral part of a language learning system, foreign-language tests are characterized by their specific functions, types, control means, and test objects, which have experiences significant transformations in the process of their historical development.

Curricula, teaching materials, and language tests of the 1960s were affected by the structural linguistics. Structuralism understood language as a system of lexical elements and grammatical structures that could be considered and classified independently. Rules of the communicative use of a foreign language in structuralism did not matter at all.

That is why test developers of 1960s-1970s proceeded from the fact that the level of foreign language proficiency could be checked only by testing with the help representative patterns consisting of all lexical elements and grammatical structures. While compiling such patterns, they relied on the studies of contrastive linguistics. Thus, vocabulary and grammar, being of special difficulty for students because of their mother tongue, were tested first of all.

One more typical thing for language tests of 1960s-70s was the fact that they did not pay much attention of the objectivity of test assessment. That is why the tests of



that time (even those, aimed at checking skills of written and oral speech) had practically only isolated, receptive, decontextualized tasks in the multiple-choice format. Consequently, productive skills could not be tested with the help of tasks like that since such skills could not be assessed separately – they can be evaluated only by assessing their corresponding partial skills. For instance, criteria of assessing the skills of oral speech include also the ones used for evaluating pronunciation and intonation.

Such a structural-analytical approach to the test development was used up to 1970s. At the late 1970s linguistics demonstrated a turn from structuralism to pragmatics and sociolinguistics; as a result, context-based use of language was put into the centre of testology. Language is considered as a communicative act in different contextual situations.

Thus, models of communicative competence from the late 1970s turned to be the important point of didactics and methodology of the development of language tests. These models can be divided into two groups: descriptive models and process-oriented ones. Descriptive models are characterized by the fact that they describe all the components of knowledge and skills required for effective communication in a foreign language. Descriptive models include also Common European Framework of Reference for Languages. Process-oriented models try to show that all the components of communicative competence are psychologically interconnected forming a union of independent factors.

Such a control method as testing can be applied successfully for current, intermediate, and final knowledge assessment. Tests help check the knowledge of students both selectively and in general. Ye.G. Zharinova and Ye.V. Cherniaieva highlight that testing is important for knowledge consolidation, improvement of the obtained knowledge and skills, and identification of knowledge gaps. Students of higher educational institutions are more scrupulous and responsible in testing their knowledge as their studying is characterized by much higher requirements for their mental activity and independency. To adopt the requirements for the curriculum they study, there should be quite high level of development of generalizing, logical, and abstract thinking since students of older ages are more intended as for their studying and perceive it as something suitable and useful for their life tasks and future profession. Therefore, testing is a method for checking knowledge and skills where a testee performs certain set of special tasks [1].

Foreign language belongs to the academic subjects, which control and mastering cannot be limited only by the written form. Adequate and comprehensive assessment of the students' knowledge in foreign languages requires a complex of tasks, which performance means active use of the acquired skills both in terms of written and oral speech.

Special attention should be paid to the problem of development of the criteria for proper assessment of language test results. According to V.B. Balabanov and T.N. Balabanova, there is no uniform strategy in this issue. Depending on the number of tasks, their form and degree of difficulty, maximum mark is set for the perfectly completed tests. Thus, while developing a control test, one should mean several levels of complexity according to different types of speech activities for a



comprehensive analysis of the student's mastering the language material meant by the programme [2].

Along with the language model, such qualities of tests as validity, reliability, and objectivity are also of great importance during the test development. These criteria require that the tests assess students' progress as accurately as possible (content validity), provide maximum reliable check (reliability) and highest objective evaluation (objectivity of test procedure and assessment).

Content validity. Validity is one of the most important criteria of test quality meaning test appropriacy for measuring the issues it is aimed to measure, i.e. correspondence of test methodology to the concept being measured. Validity is determined by correlation of the testing results with the success of completion of the corresponding activity. Validity is considered quite high if the correlation coefficient is more than 0.6. That is not as simple as it sounds since this criterion requires exact description of the test purpose and thorough check of the compliance of test purpose with the test tasks.

One can single out various components of a general concept of "validity". Content validity of a test means that the test tasks cover completely (and in the necessary proportion) all the key aspects of that knowledge area, which preparation is evaluated by the test immediately. That is why the condition of the development of content-valid tests to check receptive skills is the specification of the test purposes that demonstrates clearly, which text types should be used. For instance, if a reading test should evaluate understanding of the main idea of a text and separate statements, then global and detailed reading should be tested. If a test is to check how the candidates can find specific information, the tasks should check selective reading [3].

Items of reading or listening tests should be oriented to real situations of the foreign language use. Therefore, the basis of such test tasks should be represented by authentic or slightly adapted texts. Besides, all the texts are read and listened to differently. That depends on the facts to be learnt from the texts, i.e. on the purpose of reading or listening. For example, train schedules are not usually read in detail from beginning to end; usually, selective reading is applied to find some specific information (time, destination etc.). On the contrary, while reading a recipe of some new dish, detailed understanding of the text is required.

Reliability. It is one of the important quality criteria belonging to the accuracy of psychological measurements. Reliability reflects a degree of accuracy and constancy, with which personal quality is measured, and characterizes freedom from errors of a testing procedure. The higher the test reliability is, the fewer errors the test contains. If possible, a test should check the knowledge as accurate as a measuring tape that shows the same results while measuring one and the same objects for several times. In this context, precise formulation of tasks, being often a cause of inexact measurement, is of the greatest importance. If a task is formulated not exactly or in a too complicated way, then the testees perform the task differently as it is not clear for them what they should do. However, the tasks performed differently cannot be assessed with a sufficient degree of reliability. Thus, test developers should be aware of formulating the tasks as clear as possible so that they can be done unambiguously.



Reliability is tested statistically after the moment when the test is approved by the sufficient number of candidates (150-200 people), having corresponding level of knowledge; it is considered as the stability of results while repeated testing in terms of the same sampling of the testees. The repeated testing is usually organized two weeks after the first testing. It is believed that in case of coefficient of correlation being 0.75 the reliability level is quite appropriate.

Reliability also considers the issue on the test length, i.e. how many tasks it should include. Reading and listening tests are to contain minimum 20 questions to measure the correctness of test performance.

Test reliability and validity are interconnected. Reliability has the sense only if a test is characterized by the sufficient validity. If a test has no reliability, measuring the quality inaccurately and unstably, then its validity is brought into question.

Objectivity. We can talk about the test objectivity only when the test results are not dependant on a tester, i.e. if different testers reach similar results while measuring one and the same parameter. In other words, a test is considered objective if one manage exclude completely intersubjective influence of a tester.

On the one hand, this criterion concerns the condition of testing and means that all the candidates are tested in terms of equal previously specified conditions, i.e. there should be determined exactly how much time it will need for task performance, how many points will be given for each task, and if the auxiliary means are allowed to be used. On the other hand, that concerns the assessment of test performance. Multiple-choice tasks such as texts for reading and listening can be assessed objectively while speaking and writing tasks can be evaluated in a biased way. However, development of clear criteria for test assessment can reduce considerably the possible subjectivity during the evaluation.

The important thing during the test development is the fact that test items should not require considerable time for their completion. The obligatory criteria are as follows: definitions should be unambiguous and task interpretations should not be loose. The questions should exclude the probability of formulating multiple-meaning answers. If the tasks are short and motivating for the short answers as well, that activates logical thinking of a student as there is no need to concentrate on some details. Besides, the tests should be appropriate for the quick statistic processing of the results. One more important characteristic of tests is considered to be their standartization, i.e. their applicability for the wide practical use aimed at specifying the level of language material mastering by a wider range of students who are gaining similar volume of knowledge on the basis of typical educational programmes [2].

In general, use of testing is an optimal method of the organization of classroom and independent work of students that favours rather quick identification of current problems in digesting programme-based subjects aimed at timely elimination of the detected gaps. Therefore, testing as a form to control mastering of language material and as a tool of quantitative assessment of academic success has proved to be rather effective and viable methodological technique in terms of meeting all the requirements of current methodical science but with the allowance for the individual specificity of academic process.



References

1. Zharinova, Ye.G., Cherniaieva, Ye.V. Use of testing and test items during the tests at English classes // Scientific and practical conference in the framework of “Week of Science” of FFL MAI-NIU devoted to the 55th anniversary of Yu. Gagarin flight. Collection of reports. Issue #8. M.: Pero, 2016.
2. Balabanov, V.B., Balabanova, T.N. On some features of using test tasks to control knowledge of foreign languages // Humanitarian scientific studies. 2018. № 1 [E-resource]. URL: <https://human.snauka.ru/2018/01/24780>.
3. Handbuch zur Entwicklung und Durchführung von Sprachtests. - Herausgegeben von der telc GmbH, Frankfurt am Main, 1. Auflage - 2012.

Стаття відправлена: 23.07.2022 г.

© Яременко І.А.