UDC 37.09

ACTUAL METHODS OF COMPUTER TESTING IN EDUCATION

Leshchevich E.I.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Kamlach P.V. – PhD Tech. Sc., assistant professor, associate professor of the department of electronic technology and engineering

Annotation. The aim of this paper is to describe optimal ways of computing testing. Tests have a number of advantages: objectivity and quality, scientifically based quality criteria, individualization of control, evaluating everyone according to common criteria, typification of the testing procedure and standardization of checking the quality indicators of tasks and tests. Despite the benefits of testing, there is a need to develop strategies and ways to combat falsification, as well as cheating during knowledge control. As part of solving this problem, it was decided to develop a methodology determining the reliability of testing based on stereotypical behavioral responses.

Keywords: reliability, testing, correlation

Introduction. Begin with that computer testing is widely used as a means pedagogical measurement, for to determine the level of competencies and the degree of formation of skills necessary to perform a particular activity. The quality and reliability of the results obtained are determined by testing technologies, which have become the subject of active scientific research in recent decades. Until the second half of the last century, tests were built on the basis of the classical testing theory), which is based on the theory of measurement error, borrowed from physics: it was assumed that the measured characteristics have some "true" values, distorted by random and systematic errors.

Main part. **Computer Testing Methods.**

Testing with Parallel Forms. Most popular pedagogical tests have parallel forms. In fact, the same content is revealed in them, but the questions are outwardly different. To justify the parallelism of the tests, in addition to the semantic and logical connection, a condition is introduced for the need for the equality of their arithmetic means and variances [1]. Despite the fact that the same condition is retained to justify the parallelism of individual questions. Students are first given one form of the test, and after a while - another. Especially the results are correlated, the resulting coefficient indicates the stability in the knowledge of students.

Separate Correlation. This method allows you to check the reliability in a single test [2]. The entire test, consisting of tasks or statements, is divided in half. The usual procedure for this is: the results of the subjects are added up separately in even tasks and in odd ones; the amounts obtained in halves are correlated. In order to find out what the reliability coefficient of the test as a whole is, the value of the coefficient obtained from the two halves is set using the Spearman-Brown formula 1:

$$r = \frac{k}{k-1} \cdot \left[1 - \frac{\sum p_i g_i}{St^2}\right], \qquad (1)$$

where k – number of tasks in the test,

$\sum p_i g_i$ – the sum of the variances of individual tasks [1].

The resulting correlation coefficient indicates the reliability of the test, depending on the reliability of each individual item, so it is better to call it the internal consistency coefficient of the test.

The reliability of the test is higher, the smaller the variance of the error component of the measurement and the greater the total variance. Therefore, a reliable test better distinguishes between strong and weak learners: their grades should be more differentiated. It should be noted that the currently existing system of grading in universities and schools from two points to five has the main drawback of precisely a small scope and, accordingly, has a low differentiating ability. The resulting

correlation coefficient indicates the reliability of the test, depending on the reliability of each individual item, so it is better to call it the internal consistency coefficient of the test. The reliability of the test is higher, the smaller the variance of the error component of the measurement and the greater the total variance [3].

Therefore, a reliable test better distinguishes between strong and weak learners: their grades should be more differentiated. It should be noted that the currently existing system of grading in universities and schools from two points to five has the main drawback of precisely a small scope and, accordingly, has a low differentiating ability. Another reason for the lack of test reliability is usually a small number of tasks. In test theory, the relationship between the number of equivalent tasks and reliability is known. This is the basis for another Spearman-Brown formula 2, which made it possible to estimate how the reliability of the test will increase with an increase in the number of tasks by n times:

$$r = \frac{nr}{1+(n-1)r}, \tag{2}$$

where r − is the increased reliability of the extended test;

n − the multiplicity of the increase in the number of tasks in the test.

Re-testing. Moreover, the same test is given in the same group of subjects twice, after a certain period of time. The results of the first and second tests are correlated, the resulting coefficient indicates the reliability, in this case, the reproducibility of the results of the subjects. Changes in students' knowledge (forgetting, learning, etc.) over time, which affect the reliability (stability) of the results, were considered here as erroneous components [3]. The reliability coefficient obtained by the method of repeated testing is more correctly called the stability coefficient, or the coefficient of stability, reproducibility.

Proctoring. This is a procedure for monitoring the progress of a remote test (in English, "proctor" is an observer at exams at a university) [3]. This concept is already quite well known not only to experts. Researchers identify the following proctoring systems. Passive monitoring of software on examinees' computers. It is done by monitoring the applications that students are using on their PCs and switching to other services during the exam. Active software restriction on student computers. For example, the Browser Lock app blocks access to other apps during exams. Passive video surveillance using software that accesses the students' webcam to capture all activities directly [1].

So, online control takes place without an examiner, without human participation. Video and sound are recorded, suspicious human behavior is automatically detected and violations are recorded. For gross non-compliance with the rules (replacing the examinee, copying tasks), the program automatically blocks access to the system. Active video surveillance. It is implemented similarly to passive, but with the addition of real-time monitoring. One proctor (observer) can monitor several students at the same time. Methods for evaluating the reliability of computer testing are carried out on the basis of stereotyped behavioral reactions - patterns. This is a specific set, a pattern of behavioral responses or sequences of stereotypical actions of a person, in relation to any area where a person applies patterns.

In conclusion. This article discusses the basic methods of computer testing. These methods are applied to various academic disciplines. The degree of reliability of the results of computer testing is high and is determined using mathematical statistics, sampling the correctness of answers.

References

1. *Methods of using information and communication technologies in the educational process: a study guide. Part 1: Conceptual foundations of computer didactics / Ural. state ped. un-t, Institute of Informatics and Inform. Technologies, Dept. new inform. technologies in education. - Yekaterinburg: [b. i.], 2013.*

2. *Mayorov, A. N. Theory and practice of creating tests for the education system / A. N. Mayorov. - M. : Intellect-center, 2001. - 296 p.*

3. *Kim, V. S. Testing educational achievements. Monograph / V. S. Kim. - Ussuriysk: UGPI Publishing House, 2007. - 214 p.*