

DIDACTICAL ASPECTS OF TEST CREATION: THEORETICAL COMPONENT

Zoya Voronova, Irina Holter, Maryna Romaniukha

Dniprovsk State Technical University (Kamianske, Ukraine)

Kamianske, Dniprobudivska st.2,

zoya.voronova2@gmail.com, rinulia@ukr.net, romanuks@ukr.net

***Abstract:** The article deals with the pressing problem of test evaluation of students. The paper analyzes advantages and disadvantages of students' knowledge test evaluation. Test evaluation of knowledge allows to create effective models of student assessment, using current approaches and facing challenges of studying foreign languages. Test control is used to set the level of achievement of academic groups and individual students, to analyze various forms and methods of teaching and hold final evaluation.*

Keywords: control of knowledge, supervisory and reference computer training programs, test, classification of tests, test signs, test criteria, types of test questions.

INTRODUCTION

The introduction of a new system of managing higher educational institutions and stricter requirements for the quality of student training demand to search for more effective means of evaluation. Besides, optimal organization and monitoring of the quality of educational process is of great importance too. Tests, ratings, modern means of pedagogical control and evaluation can improve both the quality of student's training and the teaching, methodological, educational activities of teachers and management activities of the administration of higher educational institutions.

In the conditions of a significant reduction of class hours it is necessary to introduce new methods for managing educational and cognitive activities of students. The most characteristic type of teaching process control is feedback management. In order to manage feedback professors need efficient evaluation methods, which permit to see if all the planned activities were successfully fulfilled. For the implementation of effective feedback, it is vital to hold regular check and evaluation. Realization of these requirements in practice can be achieved only with the help of tests that are not very time-consuming both for the students and for the teachers.

Relevance of the problem. The method of testing with the use of ICT as one of the methods of pedagogical control is one of the most scientifically grounded among means of knowledge evaluation of students. It is widely used all over the world. Currently, standardized testing technologies are distributed in the educational system of Ukraine. The theory of educational evaluation and pedagogical testing technology are being widely introduced into higher education practices.

Previous research. As the analysis of scientific works on the research problem shows, the question of evaluating students' progress in terms of implementing test control of knowledge is one of the most controversial. To date there are many different approaches, both general and subject-oriented, which take into account the specifics of individual disciplines.

Many scientists, such as E. Baker, R. Henig, V.Reddy, B. Bitinas, V. Bespalko, A. Kokchetov, V. Maksimov studied the methodology of knowledge assessment and argued its efficiency. Of particular interest are papers devoted to knowledge assessment by D. Steiner, M. Kolen, R. Brennan, G. Fulcher., F. Davidson, V. Avanesov, N. Volkova, K. Korsak, I. Pidlasiy, N. Talizina, H. Tsehmistrova, M. Chelishkova, whose contribution has shaped the methodological foundations of our study.

The purpose of this article is to argue the approaches to organizing test assessment of students' achievement in the conditions of credit-modular study of philological disciplines that are in accordance with the provisions of the current legal documents and are oriented to take into account the nature of the educational and cognitive activity of future teachers and interpreters.

Uniqueness of the issue. The goal of present research is to investigate various types of test formats and ways of testing, in order to see how the theory could be applied in practice. The newness of the research is that it generated and summarized recent practice in the branch of testing. This article attempts to look at the importance of classroom assessment and evaluation advantages. The methodology of this paper is a descriptive approach, using classroom experience and relying on other research carried out in this regard. Classroom assessment and evaluation are highly concerned with qualitative judgments that are used to improve students' knowledge and learning. Assessment and evaluation also give teachers useful information about how to improve their teaching methods. Evaluation goes beyond students' achievements and language assessment to consider all aspects of teaching and learning.

1. THE ROLE OF COMPUTER TESTING IN EDUCATIONAL INSTITUTIONS: THEORETICAL REVIEW

Much work is being done in many countries to expand the scope of computer testing in educational institutions. For example, several states in the United States have introduced or are planning to develop and implement computer testing in response

to federal government testing requirements to assess the performance of educational institutions and teachers (Fetisov 2011; Bateshov 2011).

Based on this, testing in pedagogical science is considered as a method of pedagogical diagnostics, besides, the academic community study the differences of the test from other forms of assessment, different approaches to the classification of tests are determined. The most common is the classification proposed by Norman E. Gronlund, who distinguishes the incoming testing; formative and diagnostic testing; thematic, final, cross testing and evaluation of residual knowledge. This classification is also respected by Russian researchers V. Zvonnikov, M. Cheleshkova, A. Mayorov, and others.

There has been a lot of research in the field of testing all over the world over the last 10 years. M. Kolen and R. Brennan (Kolen, Brennan 2014) consider the role of tests very useful and important, especially in language learning. It is a means to show both the students and the teacher how much the learners have learnt during a course. Moreover, according to them, tests could be used to display the strong points and weaknesses of the teaching process and help the teacher improve it. They can demonstrate what should be paid more attention to, worked out and practised. Furthermore, the tests results will display the students their weak points, and if carefully guided by the teacher, the students will be even able to take some remedial actions.

R. Henig (Henig 2013) believes that students learn more when they have tests. According to his works, tests decrease the practice and instruction time. What he means is that the students are as if limited; they are exposed to practice of a new material, however, very often the time implied for it is strictly recommended and observed by a syllabus. That denotes that there will be certain requirements when to use a test. Thus, the students find themselves in definite frames that the teacher will employ. Nevertheless, there could be advantages that tests can offer: they increase learning, for the students are supposed to study harder during the preparation time before a test. He emphasises the idea that the learners study harder for the classes where they are tested thoroughly and claims that the students want and expect to be tested.

L. Shepard (Shepard 2013) thinks that too much testing could be disastrous. It can entirely change the students' attitude towards learning the language, especially if the results are usually dissatisfying and decrease their motivation towards learning and the subject in general.

D. Steiner (Steiner 2013) assumes we should not forget that the tests when administered receive less support from the teacher as it is usually during the exercises in a usual language classroom. The students have to cope themselves; they cannot rely on teacher's help if they are in doubt. During a usual procedure when doing various activities the students know they can encounter the teacher's help if they require it. They know the teacher is always around and ready to assist, therefore, no one is afraid to make a mistake and try to take a chance to do the exercises. However,

when writing a test and being left alone to deal with the test activities, the students panic and forget everything they knew before.

G. Fulcher and F. Davidson (Fulcher, Davidson 2007) think that students' encouragement is a vital element in language learning. Another question that may emerge here is how to reach the goal described above, how to encourage the students.

So, we can speak about the tests as a tool to increase motivation. However, having failed for considerable number of times, the student would definitely oppose the previous statement. Hence, we can speak about assessment and evaluation as means for increasing the students' motivation

Besides, there are works that touch upon computer tests and their role in teaching foreign languages: J. D. Brown (Brown 2008), T. L. Chiu, H.C. Liou, and Y. Yeh (Chiu, Liou, Yeh 2007), T.F. McNamara and C. Roever (McNamara, Roever 2006), M. Takimoto (Takimoto 2009), K. Zechner, D. Higgins, X. Xi, D.M. Williamson (Zechner, Higgins, Xi, Williamson 2009)

The main characteristics of tests as defined by modern testing science are the following: validity (conformity of the used materials to the objectives of assessment); reliability (continuity of results of test control with repeated use); comprehensiveness (the scope of study material presented in the test); standardization (setting the single procedure for conducting and summing up the results of testing) (Shymkova 2007; Baker 2013; Kane 2013).

The separate kind of tests are multimedia tests – they are computer testing programs created using a whole arsenal of multimedia, with a combination of text material, audio and video information, photographs, painting reproductions, animations, and the like.

The analysis of scientific and pedagogical literature (V. Avanesov, V. Kuznetsov, A. Mayorov, Z.Slepkan, etc.) showed that the basic conditions necessary to hold test evaluation are:

- taking into account the classical and modern test theory: on the basis of test theory and modern methods of test design it is possible to provide reliability, validity and efficiency of assessment; it is also important that test control is not limited to evaluating student knowledge; tests help assess the entire students' educational activity, in particular the dynamics of their overall development, the formation of special skills and abilities, activity, cognitive interests and creative abilities;

- professional interest and creative attitude of the teacher to the organization and management of educational process. According to V. Avanesov, "tests can be effective only in such an educational process where the teacher ... becomes a developer of new software and methodical tools, the organizer of the process of self-education of students. Training should begin with the entrance test evaluation, be accompanied by self-assessment and end with a final test" (Avanesov 1998);

- active participation of students in the organization and implementation of their own educational and cognitive activities within the learning process by self-assessing the results of their learning.

In comparison to traditional forms of assessment, computer testing has several advantages:

- prompt receipt of test results and freeing the teacher from the labour-intensive work on the processing of test results;
- objectivity of assessment;
- computer testing is more captivating compared to traditional forms of polling which creates positive motivation for students;
- improving the efficiency of the teaching staff (Avanesov 1998; Henig 2013; Fetisov 2011; Kuznetsov 1999; Shepard 2013; Shymkova 2007).

However, despite some achievements, the problem of diagnosing the academic achievements of future philologists through testing has not yet received a thorough study and scientific synthesis.

We see a pedagogical test as a system of tasks of a specific form, interrelated by the course material with increasing complexity, promoting to evaluate the structure and measure the scope of knowledge and other characteristics of the individual.

Typically, a test is a standardized method for determining the level and structure of students' knowledge. In this kind of test, all participants receive the same tasks, under the same conditions and with the same rules of assessment of responses. This method allows us to range participants according to the level of their knowledge and on this basis, to objectively determine the place in the group (or rating) of each student.

Unification and standardization of the structural representation of test tasks is especially needed to let the students understand the instructions fully, to objectively evaluate the test results, to use ICTs to hold the assessment and analyze the test tasks.

The form, type and kind of test tasks affect their structure, the principles of content design, determine the level of skills and professional training of students. Therefore, it is necessary to rely on some general principles for constructing test tasks, regardless of the type of specific discipline they will be created for (Mereshchuk, Barkhayev, Stadnik 2006; Steiner 2013).

The test task is an integral part of the pedagogical test, which meets the requirements of technological capacity, form and, moreover, statistical requirements: the specified complexity.

Test tasks can be divided into two groups: closed - tasks with a variety of options; tasks with alternative options; the tasks with multiple choice; tasks for matching correspondences; tasks for establishing the correct sequence and open ones - the open tasks; add-on tasks.

It should be noted that not all types of tests allow you to use a computerized version of programmable assessment. Let us define them. An open task type, or, simply speaking, keyboard input - is a very powerful tool when checking various kinds of terms, constants, dates. However, its implementation, as a rule, is mathematically complex and therefore most of the developers skip them. The problem is, first of all, that the phrase entered must be subjected to a syntactic, and ideally, semantic analysis that simulates the variants of the possible text responses of the person who answers the test. In addition, the student can make an automatic mistake and in most areas of knowledge, such machine errors can't be considered an error, which requires a very flexible implementation of computer logic (Avanesov 1998; Koh, Reddy, Chatterji 2014; Kordon 2012; Kuznetsov 1999). Besides, we need to mention that a language learner can enter various synonyms that may not be provided by the database developer and at the same time can be absolutely or partially correct.

It is implied that there may be several possible answers in an open question. There are also a number of variations to open questions.

1. Enter multiple responses in a certain sequence. It can be used in questions demanding the narration of strict sequence of any operation. The type of question is as difficult to program an open one, it is complex in designing and causes certain difficulties for students, since it requires not only error-free input of answers, but also an error-free reciprocal arrangement. However, despite its rather rare application, this type is indispensable and is a powerful means of determining the level of knowledge of students in matters such as the sequence of the transformation of matter in chemistry, the sequence of actions of various types of repairs, etc.

2. The selective type of question. This is a classic variant, which most developers consider necessary and sufficient for computer testing. In this type of question, one or more correct answers from the proposed ones can be taken into consideration. Some theorists draw a dividing line between these two varieties of questions, but from the point of view of formal logic, these varieties are absolutely equivalent. Computer implementation of this type is incredibly simple. Perhaps it is precisely what makes it a popular choice in various existing testing programs. For the implementation of this type, it is sufficient to have basic knowledge in any programming language or in programmable office systems such as Excel or Quattro.

In the selective type of question, there are also varieties.

1. An alternative type is the simplest form and assumes a ready-made answer already in the text of the question. It is only possible to specify whether the answer is correct or not (answer "Yes or No"). In spite of its simplicity, this type can be successfully used in a few areas of knowledge.

2. Consecutive type of question. It is the most difficult for students, although quite simple in implementation, it gives the teacher a powerful tool for evaluating not only specific knowledge but also logic.

The open or closed form of test tasks also determines the level of development of students' mental activity. This requires constructing tests that use educational elements that develop three levels of cognitive-mental activity - the level of analysis and synthesis, algorithmic and intellectual-search levels. A variety of test tasks provides a depth of perception of the content of discipline provided that we wisely select the elements of the discipline to diagnose the level of knowledge through testing. Undoubtedly, it is necessary to compile several test-tasks of various types for each study topic in order to form a bank of test tasks of the corresponding course.

The development of test-tasks and tests as means of diagnosing the level of educational-professional training begins with the division of skills and knowledge, identified by the educational-qualification standards of philological degree. The required scope of foreign language skills can be represented by the modules of the disciplines, which are given in the curriculum of training. In turn, these modules are formed with the help of learning blocks. It is the mastery of these learning blocks which is evaluated by the testing of appropriate skills – those of structuring new ideas, conceptual-analytical and productive-synthetic ones. Therefore, the learning blocks must be clear-cut components of the curriculum, structured in a modular way (Baker 2013; Kane 2013; On conducting a pedagogical experiment 2003; Steiner 2013).

Systematic objective assessment, especially its ICT-based version is also an effective means of differentiating and individualizing the learning process based on the analysis of test results and increasing the motivation of students to systematically study independently during the school year. Students are actively involved in the organization and implementation of their own educational and cognitive activity through self-control and self-examination by testing the results of their independent work (Henig 2013; Kuznetsov 1999).

2. TECHNIQUE AND METHODOLOGY OF TEST DESIGN.

2.1. Division of the tests and their ones

According to the goal and importance in the educational process assessment tests are divided into the ones for formative and summative assessment.

Formative assessment is applied at all stages of learning and are designed to track the level of mastery of study material, fix it and recapitulate it. Their main goal is to identify gaps in students' knowledge.

Tests for summative assessment are carried out as some kind of summary of the work on the studied topics.

It is advisable to carry out tasks for formative assessment at different stages of the lesson (during brainstorming, when detecting the scope of students' perception of the study topics and during its consolidation), but not more than for 10-15 minutes.

As long as these tasks are formative, they should not be made too complicated. At this stage of assessment, it is necessary to use predominantly such test tasks as selective, sequential, or ordinal, constructive, schematic tests.

The data from thematic assessment informs the teacher about the level of students' learning material and indicates how wise the test design algorithms were. In order to ensure maximum reliability of the evaluation results and to check comprehensively how fully the topics were studied, the instructor must prepare a test with a large number of tasks. It should fully reflect the content of the topic, should give a balanced presentation of both the main theoretical material and practical problems. When preparing materials for modular control, the teacher must make them well-balanced in terms of complexity. To do this, developers should include tasks in approximately this proportion: 25% are easy tasks, 50% - average complexity, 25% - complex. This distribution will reflect the ratio of strong, average and weak students in the group.

When designing a test, one must take into account that the proposed tasks must be diverse in terms of content and form, they will not be monotonous, but will provide a permanent and steady motivation of students to work. It is not worth to limit the tasks by closed form, it is wise to offer to students open tasks too, where they have to write a complete answer to the question themselves.

The number of test tasks combined in one pedagogical test is determined by the term "Test Length" or "Test Volume". To assess students' knowledge in the process of express control (15-20 minutes) you can use a test with a maximum length of 15-20 test tasks; at modular control of knowledge and skills of students, designed for a whole lesson, it is required to include the maximum of 40-50 test tasks.

Specialists, for example, emphasize that the reliability and objectivity of the test check of knowledge and skills increases with the increase in the length (volume) of the test. In order for the tests to fulfill the above-mentioned functions, it is necessary to develop a test designed wisely from the didactic and substantive side, as well as to evaluate their compliance with the educational standard.

For the effectiveness of the thematic control of knowledge and skills of students through testing the teacher needs:

- to have at least four variants of test kits;
- to assign the number of points for correct answer to the given question for every task of the test and for the whole set of tests (the number of points in each variant must coincide);
- to have a key with variants of correct answers of test tasks in all types of tests,

2.2. Methodology of students' knowledge evaluation from test results.

Teachers can transfer the points received to 5-point assessment scale by assigning value to each task. For example, for each correctly performed task that verifies the knowledge of the theoretical material or task with the choice of one correct answer

students can get 1 point, the task with the choice of several answers or finding correspondence - 3-4 points. Having obtained the total number of points, the teacher determines the intervals that correspond to the international standards of ECTS. The following distribution of points is accepted: from 100 to 90% of the points corresponds to high level; from 89 to 60% - sufficient; from 59 to 30% - average; from 29% and below - the initial one.

So, students can be evaluated with the highest score if all tasks are correctly performed, but also if their test results are very close to the best score. Thus, the student is given some opportunity to make a mistake, which he does not have during the traditional evaluation.

If the number of points obtained for the test by more than 50% of students is below the maximum, one should pay attention to the correctness of the test task, the relevance of its content to the test material.

For proper introduction of testing as a methodical form of educational process it is mandatory to make it systematic, to ascribe a role to it in the general structure of studying and to develop methods of its use throughout the school year, not only during the course of thematic certification, but also at the lessons of studying and consolidating new material, which will be a good training for students, will help them get used to performing similar tasks. It is crucial for the students to successfully master the algorithms of performing test tasks of various types, levels of complexity and structure, to discuss test results, to identify typical mistakes and determine the ways to eliminate them. Only this approach to the implementation of test technologies will contribute to the qualitative development of their modern educational institution.

3. COMPUTER PROGRAMS FOR CREATING TESTS.

There is now quite a big market for computer testing programs. Let us discuss the most widely used and the most user-friendly ones.

"Test Designer 3.3" is a versatile knowledge test program that allows you to use an unlimited number of topics, questions and answers. The program supports five types of questions and suggests any types of tests. It is possible to support the tests with music, sounds, images and videos. Any data can be printed on the printer, exported to files of different formats (Word, Excel, Access, HTML, XML, Text file, Paradox, DBase, etc.). It also gives the opportunity to set the value of questions and answers, to limit the answers in time, assign a score after end of testing (the scale of assessments can be configured from 2 to 100-point system), check spelling. On a single computer testing can be conducted by several students independently of each other by entering the program under their own names.

"Tests designer" consists of three parts: "Editor" (designed to fill and edit the database); "Simulator" (created for testing procedure based on the database

"Editor"); "Results Administrator" (designed to analyze the test results). The program has many features in terms of combining multimedia: questions can include music, sounds, images, videos, formatted text of unlimited length. In addition, "Tests designer" supports the main types of test tasks of the closed and open type.

MyTestX is another well-known system of programs for creating and conducting computer testing, collecting and analyzing the results, evaluating test results on the scale indicated in the test (program can be downloaded for free from the address <http://mytest.klyaksa.net>).

The UniTest System Universal Test is another program to create tests, to hold testing (locally, on CDs, in conjunction with electronic tutorials, on a local network of educational institutions, etc.), as well as for a detailed analysis of test results.

The UniTest System program, as well as its predecessors, allows you to use all forms of test tasks: to choose one correct answer, select several correct answers from the proposed variants, set the match, set the sequence, enter a response from the keyboard or enter a free answer with keyword search. You can use formatting, graphics, audio and video files in the question texts. There is a system for automatic importing of already prepared test tasks from any Microsoft Word documents or plain text documents.

The UniTest System consists of 5 main modules: Editor, Test, Report, Settings and Monitor (Server and Monitor for testing in the network). In addition, the package includes UniTest Starter (Quick Launch) and UniTest Direct (Internet Recovery).

PowerPoint, which is part of the Microsoft Office Word suite, also offers a wealth of opportunities for creating multimedia tests and allows you to combine different kinds of information representation, apply hyperlinks and links.

4. OUTCOMES

It is very important to use theoretical recommendations in practice. Building up on elements of the above mentioned test platforms we developed our own program based on HTML-5. It is the latest version of Hypertext Markup Language, the code that describes web pages. It's actually three kinds of code: HTML, which provides the structure; Cascading Style Sheets (CSS), which take care of presentation; and JavaScript, which makes the things happen. The set of tests is for offline use, it allows for the use of multimedia, for instance, images, audio and video files to let the assessment reach audial and visual language learners.

Below are the types of questions with textual material that can be presented with our html-based testing program:

MULTIPLE CHOICE

Text comprehension:

THE LAWS OF HEALTH

There are certain laws of health which deserve particular attention and they are so simple that even a child can learn them. A constant supply of pure fresh air is indispensable to good health. To secure this, nothing impure should remain either within or near our homes, and every room in the house especially the bedrooms, should be properly ventilated every day. (4, c.400-401).

1. Who can learn laws of health?
a) adults b) children c) adults and children
2. What is indispensable to good health?
a) pure fresh air b) pure fresh water c) pure fresh carbon
monoxide
3. What room in the house should be properly ventilated first of all?
a) kitchen b) bedroom c) living room

Grammar test:

1. This is the task I've ever done.
A. ease B. easiest C. easier
2. There aren't slices of pie left.
A. more B. much C. a lot of
3. It was the book you had left on the bench.
A. which B. whose C. what
4. They can everything at the moment.
A. see and hear B. are seeing and hearing C. saw and heard
D. were seeing and hearing

MATCHING

Establish correspondence between the parts:

1. They were about to leave a) and luckily I was at home.
2. He called on me last Saturday b) he will soon come around.
3. Brian began on his third painting c) when you rang.

4. Don't worry about the chairman d) for the exhibition.

Match proper text to image (Figure 1):

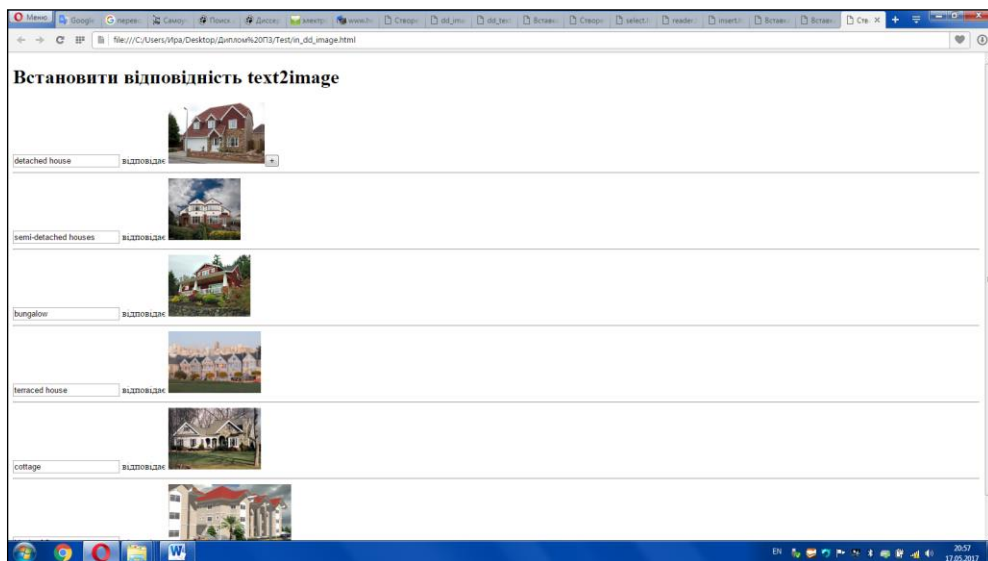


Figure 1. Results of taking the matching text-to-image test

Source: *Own Work*

GAP FILLING:

Closed-end questions:

Is losing, lost, loss, consumes, consumers, division, divide, independent, independence, independently, to depend on, faint, faintly

1. Mankinda lot of natural resources harmful for our planet.
2. Kate is so oblivious. She always..... her books and purses.
3. My cat is a very animal. She always does what she wants.
4. After that horrible he still cannot recover.
5. This large hypermarket always makes nice presents for their
6. Amanda was looking at her son and couldn't say a word.

Open-end questions. Insert proper article:

1. Come to ... blackboard and write. 2. You have ... spelling mistake in ... word "nursery". 3. He is ... old friend of mine. 4. There came ... tap at ... door and in another moment we saw ... small girl enter ... room. 5. He is ... young artist and, I should say,

rather talented. 6. He gave her ... cigarette and lighted it. 7. I don't feel ... sympathy towards this man.

TRUE/ FALSE

6. Choose the sentence which is closer in meaning

1. They were about to leave when you rang.
 - A. They left something.
 - B. They were going to leave.
2. He called on me last Saturday and luckily I was at home.
 - A. He came to visit me.
 - B. He phoned me.

CONCLUSIONS

The value of tests as a tool for evaluating the effectiveness of student learning and cognitive activity in comparison with other forms of control lies in the fact that, firstly, the tests are a much more qualitative and objective method of evaluation and, secondly, test indices are focused to measure the scope of assimilation of key concepts, topics and sections of the curriculum, skills and abilities, but not just to confirm that the student has some formally acquired knowledge (Henig 2013).

The use of test evaluation in training philologists allows teachers to develop effective models for assessing students' achievement with maximum use of modern approaches, specifics, goals and objectives of learning a foreign language, and, taking into account existing trends, opens ways to computerize test evaluation.

Among the main advantages of test evaluation of students in linguistics we can distinguish:

- taking into account the individual characteristics of students; the possibility of analyzing in detail the scope of student's knowledge in each content module of the course;
- implementation of efficient diagnostics and feedback with each student; gaining the learning time in the course by ongoing knowledge evaluation and the objectivity of evaluation of learning outcomes;
- a large selection of forms and means of evaluation;
- saving teachers' time and efforts by automating knowledge assessment.

At the same time, along with the positive ones we should mention the negative features of tests. In particular, test control does not allow the teacher to check the student's oral presentation.

Here we gave a short theoretical overview of such testing platforms as Test Designer 3.3, MyTestX, The UniTest System Universal Test and Microsoft PowerPoint.

Having launched an offline testing system for the language learners we are trying to automate the process of their knowledge assessment, looking for ways to evaluate writing skills and increasing motivation to study languages.

ACKNOWLEDGMENTS

The research leading to these results has received, within the framework of the IRNet project, funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement No: PIRSES-GA-2013-612536.

REFERENCES

- Avanesov, V.S., 1998: *Composition of test tasks: A training book for teachers of universities, teachers of schools, post-graduate students and students of teacher training Universities*, 2nd ed-n., corrected and elaborated, M.: Adept, 1998, 294 p. [In Russian]
- Baker, E.L., 2013: The chimera of validity, *Teachers College Record*, Vol. 115 No. 9, [on-line] at www.tcrecord.org (accessed 15 June 2017).
- Bateshov E.A. *Fundamentals of computerization technology development: A training book*, Astana: Poligraf-Mir, 2011, 241p. [In Russian]
- Brown, J. D., 2008: *Raters, functions, item types and the dependability of L2 pragmatics tests*. In E. AlcónSoler, & A. Martínez-Flor (Eds.), *Investigating pragmatics in foreign language learning, teaching and testing*, pp. 224-248.
- Chiu, T. L., Liou, H.C., Yeh, Y., 2007: *A study of web-based oral activities enhanced by automatic speech recognition for EFL college learning*. *Computer Assisted Language Learning*, 20, 3, pp. 209-233.
- Fetisov, V.S., 2011: *Computer technologies in testing: teaching aid*, Nizhyn, Publisher PE Lysenko M.M., 2011, 140 p. [In Ukrainian]
- Fulcher G., Davidson F., 2007: *Language Testing and Assessment*, New-York: Routledge, 396 p.
- Henig, R.J., 2013: *The politics of testing when measures 'go public'*, *Teachers College Record*, Vol. 115 No. 9, [on-line] at www.tcrecord.org.

- Kane, M., 2013: *Validity and fairness in the testing of individuals*, In: Chatterji, M. (Ed.), *Validity and Test Use: An International Dialogue on Educational Assessment, Accountability and Equity*, Emerald Group Publishing, Bingley, pp. 17-53.
- Koh N., Reddy V., Chatterji M., 2014: *Understanding validity issues surrounding test-based accountability measures in the US*, *Quality Assurance in Education*, Vol. 22 Issue: 1, pp. 42-52, available at <https://doi.org/10.1108/QAE-12-2013-0051> (accessed 1 June 2017).
- Kolen, M. J., Brennan, R. L., 2014: *Test Equating, Scaling, and Linking: Methods and Practices*. USA: Springer, 172 p.
- Kordon, Yu. V., 2012: *Test control as one of the methods of checking the knowledge, skills of students in higher medical educational institutions*. In: *Modern information technologies and innovative methods of training of specialists': methodology, theory, experience, problems: a collection of scientific works*. Iss. 30, Kyiv-Vinnitsa, 2012, P. 255-259. [In Ukrainian]
- Kuznetsov, V.G., 1999: *The concept of development of test technology to control the level of student learning in the system of vocational education in Russia*. *Testsineducation*. Scientific-methodical newsletter with electronic attachment, 1999, Iss.1, P. 30-45. (In Russian)
- Mcnamara, T.F., Roever, C., 2006: *Language Testing: The social dimension*. Oxford: Basil Blackwell, 292 p.
- Mereshchuk, M. Ya., Barkhayev, Yu.P., Stadnik, G.V., 2006: *Test assessment and rating in education: teaching aid*, Kharkiv: KHNAMG, 2006, 106 p. [In Ukrainian]
- On conducting a pedagogical experiment as to the introduction of credit-module system of organization of educational process in higher educational institutions of III-IV levels of accreditation. The decision of the Collegium of the Ministry of Education and Science of Ukraine from April 24, 2003, № 5/5-4. [In Ukrainian]
- Shepard, L.A., 2013: *Validity for what purpose?* An afterword, *Teachers College Record*, Vol. 115 No. 9, pp. 1-12.
- Shymkova, I.V., 2007: *Use of automated test control knowledge for organization of independent educational and cognitive activity of students* / Collection of scientific works. *Pedagogical sciences*. Iss. 46, Kherson: Publisher KhSU, 2007, P. 407-410. [In Ukrainian]
- Steiner, D.M., 2013: *Trusting our judgment: measurement and accountability for educational outcomes*, *Teachers College Record*, Vol. 115 No. 9, [on-line] at www.tcrecord.org.

- Takimoto, M., 2009: *Exploring the effects of input-based treatment and test on the development of learners' pragmatic proficiency*. Journal of Pragmatics 41, pp. 1029-1046.
- Zechner, K., Higgins, D., Xi, X., Williamson, D.M., 2009: *Automatic scoring of non-native spontaneous speech in tests of spoken English*. Speech Communication, 51, pp.883-895.