E-learning and STEM Education
Scientific Editor Eugenia Smyrnova-Trybulska
"E-learning", 11, Katowice-Cieszyn 2019, pp. 609-629
DOI: 10.34916/el.2019.11.38

AUTOMATED TESTING AS A LEARNING ASSESSMENT TOOL FOR UNIVERSITY STUDENTS

Olena Gulesha[1], Viktor Bahriy[2], Mykhailo Pyshnyi[3], Maryna Romaniukha[4]
Dniprovsky State Technical University, Kamianske, Dniprobudivska st. 2
[1]mishapyshnyi@rambler.ru, [2]annabagry@ukr.net, [3]mishap635@gmail.com, [4]romanuks@ukr.net

Abstract: *The paper considered and analysed aspects of modelling elements of knowledge test assessment. Learning process is a continuity of consecutive study of individual parts of the course and disciplines as a whole. Computer-based test assessment is gaining popularity among professors and educational managers. The authors discuss the benefits and drawbacks of digital test assessment. Computer testing is rationally applied to solve the tasks of formative and interim student learning assessment. The authors maintaim that the increase in the efficiency and accuracy of assessment is achieved not by increasing the number of test tasks, but by bringing the level of complexity of the questions to the student's level. Their approach is based on the overlay student model of student knowledge. For more reliable assessment of the individual student's knowledge the authors used the Sugeno fuzzy inference method.*

Keywords: higher education, knowledge assessment, adaptive test assessment, fuzzy logic

INTRODUCTION

In modern conditions, the system of higher education in Ukraine is focused on new educational technologies, which are connected with reduction of compulsory presence hours and increase of the share of independent work of students. In this regard, it is necessary to improve the learning strategy and monitoring of achievements based on a particular scope of individual knowledge, skills and abilities of the learners. In order to implement individual approach to every student we must carry out a qualitative and quantitative analysis of their knowledge and skills correctly (*Nacional'na doktrina… 2001*).

Monitoring allows establishing feedback between the teacher and the students, which improves the assessment of dynamics in the knowledge acquisition; it allows identification of the actual scope of knowledge in the required field, skills and competences of future professionals. Monitoring is also a field of professors' practical activity and a subject of theoretical studies. With the help of monitoring one can identify the advantages and disadvantages of new teaching methods, establish the relationship between the planned scope of learning and the actually achieved ones, compare the performance of different teachers.

The analysis of the obtained assessment results is a complex multifactorial correlation with a large number of variables, which often requires large investment of effort and time for statistical calculations. However, conducting such monitoring of students' knowledge, skills and abilities is an essential component of the learning process.

It is becoming increasingly important to hold pedagogical diagnostics of the results of educational activity in order to further improve the educational process. One of the tools to assess students' knowledge during the study of disciplines is a test system. Computer technologies play a vital role in the organisation of the modern educational process.

Evaluating the quality of learning using computer technology can significantly reduce the time and effort invested in analysis and it also increases the informativeness of results. Let us stress the fact that computer testing today is the most objective method of pedagogical monitoring. Therefore, the aspects of algorithmization and designing tools for computer-based test assessment of skills and abilities are of particular interest.

Particularly relevant is the development of algorithms and software aimed at automated solving of identified problems, which will increase the objectivity of knowledge assessment through the accumulation of data about student achievement and adapt the process of automated testing to the level of knowledge and psychophysiological characteristics of a particular student.

1. ANALYSIS OF RECENT RESEARCH

A lot of research works are devoted to the application of information technologies and the designing of student testing systems. Lizunov, Teslya, Zyanchurina, Nekrasov, Rizun, Taranenko suggest systematic account of the organizational problems in the development and application of computer testing technologies in high school. A group of authors (Zagrebel'nij, Shcherbina, Fedoruk, Artamonov, Kravchenko, Babij, Kravchenko) consider methods of adaptive testing of students' knowledge and practical aspects of computer-aided adaptive learning and testing. Research in the field of knowledge testing automation is aimed at: increasing the effectiveness of the hardware and software design of the test session (Cidelko, Yaremchuk, Shvedova; Veligura, Lekhcier,

Tkachenko), the analysis of experience using automated testing systems (Katerinchuk, Naumenko, Abakumova, Kovaleva).

The purpose of the paper is to provide a theoretical justification for testing as a means of improving the quality of the learning strategy and assessing student achievement, based on the level of individual knowledge, skills and abilities of the student being taught. Research objectives: theoretical literature review of automated student testing systems; description of test banks for computer-aided assessment of student's level and structure of knowledge, development of an algorithm for computer testing, with an account of student's model, the complexity of test tasks, the time to cope with it.

2. RESEARCH RESULTS

The analysis of research and personal experience in using computer-enhanced knowledge assessment shows its superiority over traditional methods (Katerinchuk, Naumenko, Abakumova, Kovaleva). The suggested approaches allow creation of a system of pedagogical testing and increased efficiency of teaching process in any course.

2.1 Drawbacks of computer-based test assessment

To date, there is quite a large number of automated testing systems that declare the ability to provide quality control of knowledge. In general, the work of modern developers emphasizes the efficiency of computer-aided adaptive testing, its effectiveness in the selection of tasks for each individual student, and obviously, saving time for the testing process itself from the student perspective (Louhab; Conejo; Tseng).

The choice of software depends on the purpose of the test. TestMaker, Moodle, Kahoot software complexes are currently in high demand and are quite attractive software products. However, when using the Moodle software system, users might report problems with obtaining high-quality technical support, so universities must have an experienced Moodle expert to deploy and maintain the system; professional enhancements may be required to extend functionality or correct software errors that occur in the system. TestMaker software might look a bit outdated in terms of interface, it also has irregular software updates. While using Kahoot! some students may have problems downloading on their mobile devices in case of poor internet connection.

As an automated form of knowledge monitoring and assessment computer testing has a number of vulnerabilities and due to its strict algorithmization it is subject to its probabilistic risks. This can manifest itself in the following:

1. The development of efficient test tools is a long and demanding process requiring high qualification and increased attention of a teacher or methodologist

who not only prepares the required number of test tasks, but also ensures their differentiation in terms of complexity;

1. It is impossible to figure out the causes and origins of deficiencies in students' knowledge of specific sections of a course, directly after the detailed test results;
2. Elaborate assessment of the thorough understanding of the subject, mastery of the style of reflection, typical of the course in question. During test assessment the student, in contrast to oral or written exam, does not have enough time for in-depth analysis of the topic, while the possibility of choosing between several suggested answers does not allow evaluation of skill of formulating their own answer and does not contribute to the development of skills to support their answers;
3. The test does not allow the teacher to check and evaluate the productive levels of knowledge associated with creativity, that is, probabilistic, abstract and methodological knowledge;
4. It does not promote the development of oral and written language of students. With random answers, the student becomes accustomed to working with ready-made wording and is not able to reproduce the absorbed knowledge in a competent language;
5. When analysing the data obtained it is necessary to take into account the probabilistic component, which distorts the results of the test and suggests a superficial representation of true knowledge and skills of students. At testing there is a chance of guessing the correct answers, and thus, there is no guarantee of solid knowledge in students. On the other hand, an accidental mistake is also possible in well-trained students, while the availability of several variants of answers, very similar to each other, often misleads some of them;
6. A possible loss of individual approach, as testing involves common rules, which all students are asked to follow. At the same time, there is a significant risk of overlooking a bright individuality of an unconventional mind;
7. Ensuring the objectivity and fairness of the test requires special measures to ensure the confidentiality of test tasks. When re-applying the test, it is desirable to make changes to the task, that is, there is a need to constantly update assignments and answers. On the other hand, designing efficient test tasks requires a preliminary empirical check and stable indicators of validity and reliability.

2.2 Proper place for computer testing in the curriculum

The above mentioned disadvantages suggest that computer testing cannot be the only one and universal way to test knowledge. This means that a reasonable combination of traditional assessment tools and test assessment techniques is required. To date, any taken technique of testing has a specified area of application and solves a limited range of tasks. Therefore, in our opinion, the final assessment in the university should remain traditional: exam, credit, course paper and graduation paper. Computer testing is rationally applied to solve the tasks of *formative* and *interim* student learning assessment.

Formative assessment is the main type of monitoring the student's acquisition of each course. Its main task is the regular management of student learning and its adjustment. It permits to get fundamental data about the progress and quality of the educational process. The data received after formative assessment is crucial for managing the learning process. The tasks of the formative assessment are basically the following:

- to reveal the scope, depth and quality of the student's perception and acquisition of the material;
- identify deficiencies in knowledge and outline ways to eliminate them;
- to find out the degree of students' motivation and their attitude to systematic work;
- to inquire into the reasons that hinder their work;
- to define the efficiency in mastering the skills of independent learning and identify the ways and means of their development;
- to stimulate students' interest in the course and their eagerness to learn.

Formative assessment is carried out on almost every lesson in the process of mastering the new study material (topics, lectures). Herewith, it is not recommended to allow significant intervals in assessment, because otherwise students will stop regularly preparing for classes and consolidating the material they have learnt. Formative assessment should take a small part of the classroom time so as not to rush into the presentation of new material and consolidation of the information absorbed. It is advisory to conduct formative assessment based on a set of control tasks, prepared by the teacher in advance and not exceeding the volume of 10 - 15 tasks. With formative assessment we achieve current progress, the learning process becomes manageable and students' cognitive activity gets stimulated.

Interim (cross-sectional, modular) assessment permits to evaluate the quality of knowledge absorbed by students in several sections or topics of a course. This assessment usually takes place several times a semester, and this allows you to test the quality of the acquired knowledge through a longer period and covers

more substantial sections of the discipline, which allows the identification of logical interactions with other sections. Interim assessment permits to check the assimilation of the acquired knowledge and skills, design individual learning scenario for each student, which permits to adapt the system of presenting new material and recapitulation units of material studied before. Accordingly, the method of assessment changes and students can be expected to get involved in independent constructive activity.

2.3 Aspects of adaptive testing

It is expedient to carry out interim assessment using adaptive test evaluation. The appropriateness of adaptive knowledge test assessment stems from the challenge of streamlining traditional testing. The main idea of this approach is that the increase in the efficiency and accuracy of assessment is achieved not by increasing the number of test tasks, but by bringing the level of complexity of the questions to the student's level. In most cases, this is usually associated with a decrease in the number of tasks, time, cost of testing and with an increase in the accuracy of the points obtained by the students after the test.

Adaptive testing methods involve the editing of the composition and sequence of the test tasks presented in the testing process on the basis of known regularities inherent in the test tasks bank and the information received from the respondent (student) during the testing.

One of the basic components of adaptive testing systems is *student model* that provides knowledge on the subject of a learning system that is needed to support decision-making when organizing a learning process to achieve learning outcomes. It contains quite comprehensive information about the student: psycho-physiological individual characteristics, level of student's knowledge, skills and abilities, ability to study, ability to perform tasks, ability to use the information received, personal characteristics and other parameters. The student model is dynamic, i.e. it changes in the course of studying the discipline, during the work with the system and therefore depends on the method of modelling the subject area. *The model of a student's level of knowledge* comes from the student model and includes types of educational and cognitive activities to acquire specific knowledge, skills and abilities.

Today, many colleges are developing their own comprehensive computerized systems designed to effectively control and evaluate students' knowledge. In 2018 the authors questioned professors and students of Dniprovsky State Technical University (Kamianske, Ukraine) about the necessity of using computer testing as an automated form of pedagogical diagnostics. 296 respondents were interviewed: 37 of them were full professors and 186 associate professors. The age structure of respondents: up to 50 years - 40%, 51 - 80 years - 60%. 86% of respondents support testing as a way to test students' knowledge, 14% oppose to its use. Such support is associated with widespread adoption and implementation of ICTs in university teaching.

According to the results of student survey the necessity of using computer testing was supported by 95% out of roughly 5000 students. As to the reasons ~~why~~ most answered that this is one of the requirements of the modern world for effective management of education of various levels.

Currently, professors at the department of Electronics are developing and implementing a test system for the students in the course Digital circuitry, for internal use.

In order to model the level of knowledge, the authors chose overlay model of a student (Popov, Lazareva 2015) – representing the student's knowledge as a subset of the model of subject domain formed by an expert (lecturer). In the model domain you can find several levels of hierarchy. A didactic unit can be considered the minimum structural unit and logically independent part of the educational material. The overlay model assumes that all knowledge in the course of training is divided into some independent parts (elements). For example, all training material consists of a set of topics L and a set of assessment tasks *TZ*, the implementation of which is intended to assess the quality of learning material. For each topic $l \in L$ there is a set of tasks $TZ_l \subset TZ$, and the following conditions are fulfilled: $\bigcup_{l \in L} TZ_l = TZ$ та $\forall l_1 \in L$, $\forall l_2 \in L\, TZ_{l_1} \cap TZ_{l_2} = \varnothing$.

Each level of the model of the subject area corresponds to the level of the student's knowledge model. A numerical attribute is assigned to each topic of the subject area which shows the student's understanding of the material on this topic. Moreover, the degree of assimilation of each of the knowledge units can be estimated by a percentage or probabilistic coefficient. The value of this attribute will be determined during the formative assessment. The current state of the learning process is, in fact, the projection of student's knowledge onto the domain model.

The overlay model stores data about student achievement for each area of knowledge of the subject domain. The results of performing a TZ set of test tasks during formative assessment permit to form a set of grades (O) for the student's success in learning the course material. Conducting test assessment on a specific topic TZ_l allows students to get an adequate assessment of its assimilation $O_{l_i} \in O$, so the overlay model allows you to identify what the student knows or doesn't know. At the initial stage, the model contains a very small amount of knowledge about the student, therefore, cannot give an adequate assessment, and therefore, cannot be used for adaptation. On that ground it was decided to use adaptive testing for the organization of interim assessment.

There are various ways to assess the scope of mastering a discipline for a certain length of time, or in general, based on data on the scope of mastering teaching

units. In the simple case, they set, by default, the parameters equal for all didactic units and the average estimation of the level of knowledge of each student is calculated as follows:

$$O_{c_i} = \left(\sum_{i=1}^{N} O_{l_i} \right) / N$$

where: *i* – task number, which is connected to didactic unit;

N – the number of tasks to assess all didactic units for a given period of time;

O_{l_i} – the result of i-task.

Using the dynamic overlay model of student knowledge allows you to determine the degree of assimilation of didactic units for a certain period of the learning process. Based on this information, students can be divided by the degree of mastering the discipline in four levels:

1. Level R_0 – "fail" or "inadequate achievement", if $O_{c_i} < 60$;
2. Level R_1 – "minimal achievement" or "satisfactory", if $60 \le O_{c_i} < 74$;
3. Level R_2 – "extensive achievement" or "good", if $74 \le O_{c_i} < 90$;
4. Level R_3 – "exceptional achievement" or "excellent", if $90 \le O_{c_i} \le 100$.

The applicability of adaptive testing methods is based mainly on the complexity of test tasks. Adaptive testing is a variant of automated testing system where the parameters of complexity and differential ability of each task are obtained in experimental way and are known in advance. Thus, before becoming a part of computer bank of test tasks, each of them undergoes empirical testing on a fairly large number of typical respondents.

The solution to this problem is possible by means of formative assessment, where the modern theory of Item Response Theory (IRT) is used for the statistical processing of the answers, the IRT being based on the Rasch model (Rasch, 1980). A prerequisite for the application of Rasch model is that all selected test tasks should be offered to each student. As a result, we get a dense matrix of test results, statistical processing, which allows you to get latent parameters of student's level of knowledge and complexity of test tasks. This approach can improve the quality of testing and eliminate random errors in the development of test tasks. Relying on the results of determining the complexity of test questions and the traditional classification of the scope of learning, one can divide all test tasks into several levels:

1. Level TZ_1 - level "satisfactory". The test tasks of this level can reveal the student's: basic knowledge of the course, general concepts, knowledge of terminology, formulas and laws. This kind of assessment tests the ability to act according to a model or a known algorithm. This is the basis, which lays foundations for the second more complex stage of testing.

2. Level TZ_2 – level "good". Testing aimed at verifying how successfully a student can make logical conclusions and operate basic concepts within standard, typical situations. Testing determines the knowledge of certain algorithms, formulas, laws (regularities) when performing standard tasks, the ability to conduct a situation analysis, use knowledge from different areas, topics. At the second level reproductive thinking is checked. Based on this type of thinking solutions are worked out.

3. Level TZ_3 – level "excellent". Testing evaluates the ability to find the optimal, rational solution in non-standard situations, in non-typical tasks. This is the most complex part of the task bank, which requires students to design new, previously unknown solution algorithms, explore possible solutions, act in a non-standard situation.

One of the important factors in the testing process is the time given to the answer and the time spent by each participant in the test. Exceeding the time to answer the test task can be considered not sufficient student's mastery of the material outlined, or as the use of third-party sources of knowledge by the student. Too little time spent by the student on the solution of the test task should signal the system either about inappropriate level of complexity of the tasks of this type for a particular student, or about the student's knowledge of the correct answer to this task in advance, or about simply guessing the correct answer. Small response time also indicates the need to revise the corresponding test task, which may be incorrectly constructed and the correct answer is in the question itself, that is, there is a tooltip. In any case, all these cases give an error in assessing the student's actual knowledge and this requires the introduction of a system of penalties, that is, the assessment should be reduced.

2.4 Fuzzy logic improves adaptive test assessment

We used the Sugeno fuzzy inference method for a more reliable assessment of the individual student's knowledge, taking into account the time spent on the answer. It is suggested to design the methodology for assessing the quality of students' knowledge using methods and means of artificial intelligence implemented in the Fuzzy Logic Toolbox package of the MatLab system in the form of Adaptive Neuro – Fuzzy Inference System (ANFIS). The ANFIS hybrid system is a combination of a neuro-fuzzy Sugeno method of deducing an artificial neural network of direct propagation with one output and multiple inputs that are fuzzy linguistic variables (Figure 1).

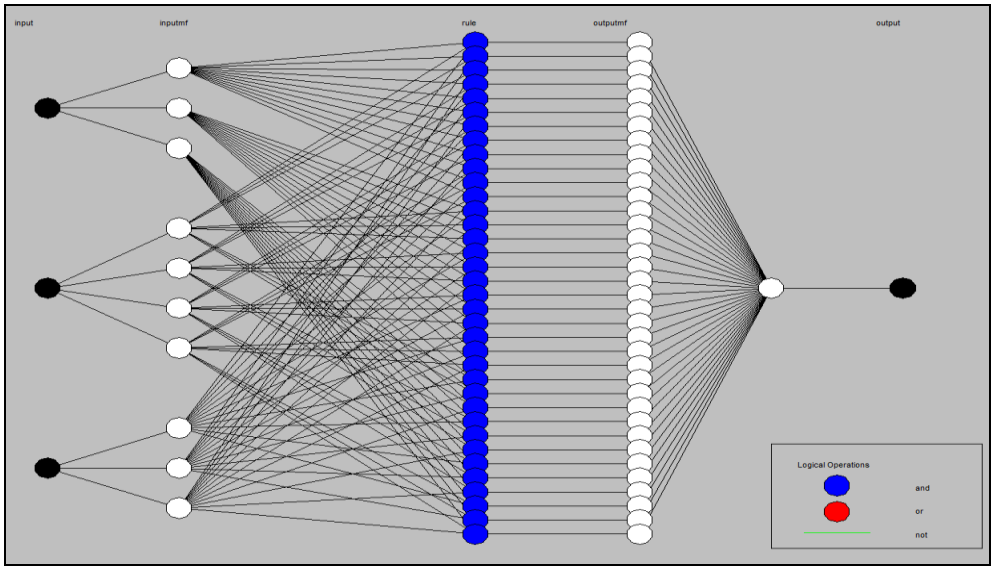

Figure 1. Neural network built in MatLab system
Source: Own work

Three fuzzy linguistic variables were considered input parameters for the system, (Figure 2):

1. "Rate of performing", which is interpreted as a term-set of values T1 = {Quickly, Norm, Slow}.

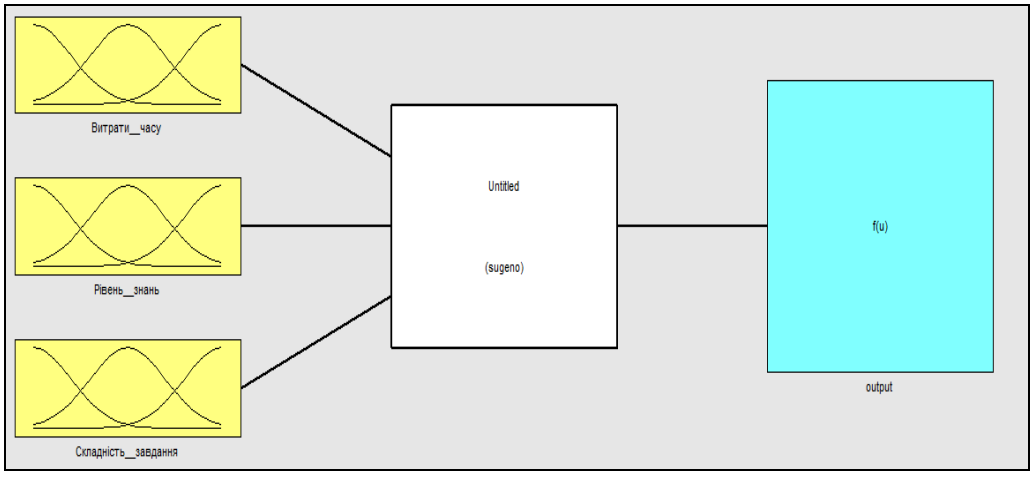

Figure 2. Input system parameters for the Sugeno method
Source: Own work

2. "Scope of knowledge", which is interpreted as a term-set of values T2 = {Poor, Satisfactory, Good, Excellent}.

3. "The level of complexity of the assignment", which we interpret as a term-set of values T3 = {Complex, Medium, Light}.

The output variable is k_s – the credibility coefficient, which we will define in the range from 0 to 1. The coefficient characterizes the decrease in the rating in case of violations of the testing procedure, for example, when the participant exceeds the time allocated for the execution of the j-task.

As a result of the analysis of the domain, a base of rules for estimating k_s is formed, which consists of two parts: input and output. The input consists of statements, connected by "I" bonds. In this case, the concluding rule is presented as follows: if (the pace of performing is fast) and (scope of knowledge - low) and (the level of complexity of the question is high) then (the coefficient of credibility - k_1) etc. The coefficients k_1, k_2, ...,$k_s \in [0,1]$ - characterize a decrease in assessment in case corresponding violations of the testing procedure take place. The surface of the fuzzy output obtained via the developed model is shown in Figure 3.

The total score, comprising the points awarded for the correct answers in the entire test, is calculated as follows:

$$O_{i\,\Sigma} = \frac{\sum_{j=1}^{m} k_{ij} a_{ij}}{a_{\Sigma\,max}} \times 100$$

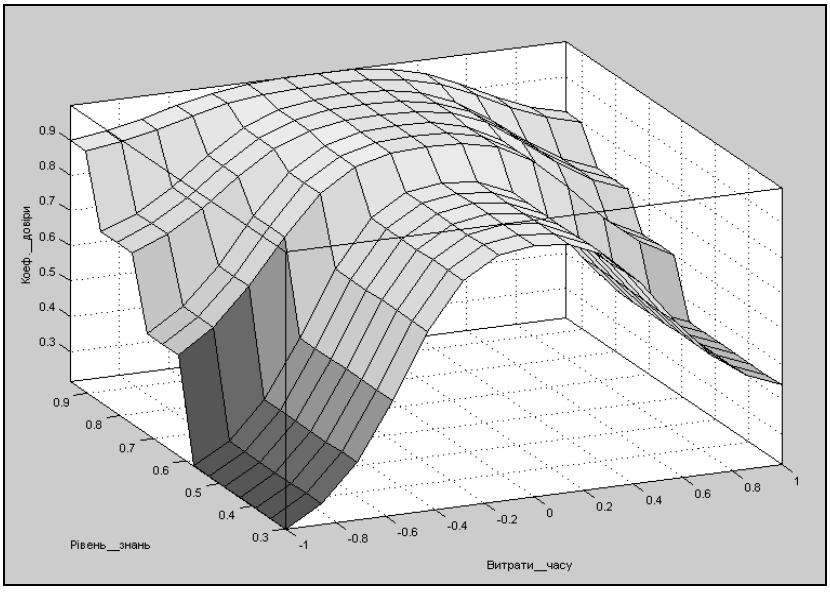

Figure 3. The surface of the fuzzy output for the value of the coefficient k_s
Source: Own work

where:

α_{ij} - the result of i-th participant for the j-th task: 1 - the answer is correct, 0 - is not true;

$a_{\Sigma max}$ - the maximum number of points for performing all test tasks;

k_{ij} is the credibility coefficient obtained through the system of fuzzy predicative rules by the i-th testing participant for each j-th test task.

2.5 Rules for efficient algorithms of adaptive test assessment

Based on analysis of results of the conducted research, we developed an algorithm of test control with block adaptation, where the decision to change tasks is made after analysing the results of testing in the previous block of tasks. The scenario of performing adaptive test is based on the following rules:

1. Using the database, the testing module carries out the identification of registered or new users, the selection of the test, tests the student by displaying the task on the screen and expecting the solution, processes the received data and records the results of testing to the database to allow further analysis and use by the teacher.

2. At first, the system puts a block of test questions of the lowest complexity (10-15 questions) to all respondents.

3. The respondent who has knowledge above the current level of complexity has a chance of early transition to a more complex level. The transition is carried out under the following conditions:

- correct answer to the first three questions of the test;
- correct answer to four questions out of the first five questions.

4. The respondent whose scope of knowledge is lower than the complexity of the tasks can complete the test ahead of time if the answers to the three questions in a row are incorrect.

5. A respondent whose scope of knowledge matches the complexity of the questions takes the whole test given, and if he or she produces 80% or more correct answers, then the system transfers the student to a more complex level.

6. The test assessment is completed on the following conditions:

- all questions in the bank of test tasks are processed;
- the scope of knowledge is assessed with sufficient accuracy;
- the student reached the end of the test;
- the respondent shows his or her incapacity to process the test questions.

Here are some of the developed test tasks for the course Digital Circuit Technology at the Department of Electronics of DSTU. The automated testing system is based on the client-server principle. With this approach, you can use all parts of the system from a variety of devices, starting with phones and ending with home computers. To use the system on your device, you need a Web browser program and Internet connection. When designing the system architecture, the principle of modularity was used for greater elasticity and convenient completion of the project in the future without the need to edit existing working structures, models and controllers. Taking into account the server approach chosen, the list of tasks and requirements was chosen by the system of storage of MySQL information in the language of SQL queries.

For the implementation of automated testing system, the best choice for the system platform is the Laravel framework. Laravel supports working with the selected database type and is created in PHP programming language. Laravel is a free, open source PHP framework.

An important feature of the test tasks in this course is rich use of non-verbal information, namely graphic information in the form of diagrams, graphs, waveforms, diagrams and more. Such information is used not only in the formulation of test tasks but also in the formulation of variants of answers. Using graphical information allows:

- stimulating higher level cognitive processes;
- making testing less tedious, more varied;
- reducing the number of accidental errors;
- arousing interest in the task and is an additional motivation in completing the test task.

Most often, they use tests where students have to choose one correct answer from the proposed answer options. These tasks consist of the following mandatory parts: 1) instructions; 2) the substantive part of the task; 3) a certain number of proposed answers (answer options); 4) reference answers; 5) evaluation systems. An example is shown in Figure 4.

The structure of a simple test task includes a question and several (at least two) answer options (one is sure to be true and the other is definitely not). A "key" is always added to the task - the correct answer (standard), which compares the answer and concludes that the task is done correctly. If the respondent's answer matches the standard, the statement becomes true, if not - the statement is false.

The purpose of the closed-circuit vehicle is to assess the ability to form a system that is relevant for testing from the proposed sign concepts. Probability of guessing the answer in the test with the choice of one correct answer:

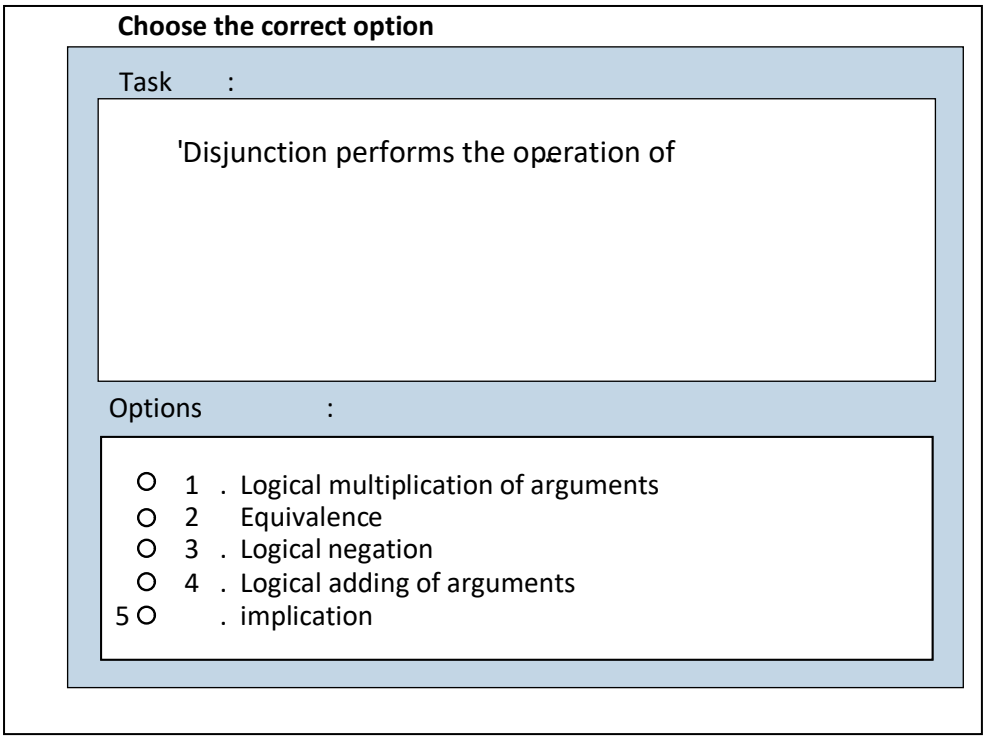

Figure 4. A test task with the choice of one correct answer from the suggested options
Source: Own work

$$p = \frac{1}{n}$$

where n is the total number of answer choices.

In tests of this type, n is 4 or 5; thus, the probability of guessing is 20 - 25%. Of course, this is a large value, so this form is used only in the simplest test tasks of the first level of difficulty, designed to test knowledge of any mandatory provisions (rules, laws, etc.). It takes 10-30 seconds to complete a simple closed-form test with one element selected.

The test of this form is most prone to the risk of cheating, clues, and other undesirable phenomena that distort the objective picture of students' knowledge in the context of even a well-organized testing process. Therefore, the developer is faced with the problem of creating several variants of tests with similar content and the same complexity.

The problem is solved by the development of 5 - 8 parallel variants of the test, for which you can use facet tasks. The facet is understood to be a form that provides the presentation of several variants of the same element of the content

of the vehicle. In addition to single-facet tasks, there can be two, three or more facets.

Each student receives only one option from the facet. In this case, all test groups perform the same test, but with different elements of the facet and, accordingly, with different answers (Figure 5). Thus, two problems are solved simultaneously: eliminating the possibility of cheating and creating tests with the same characteristics.

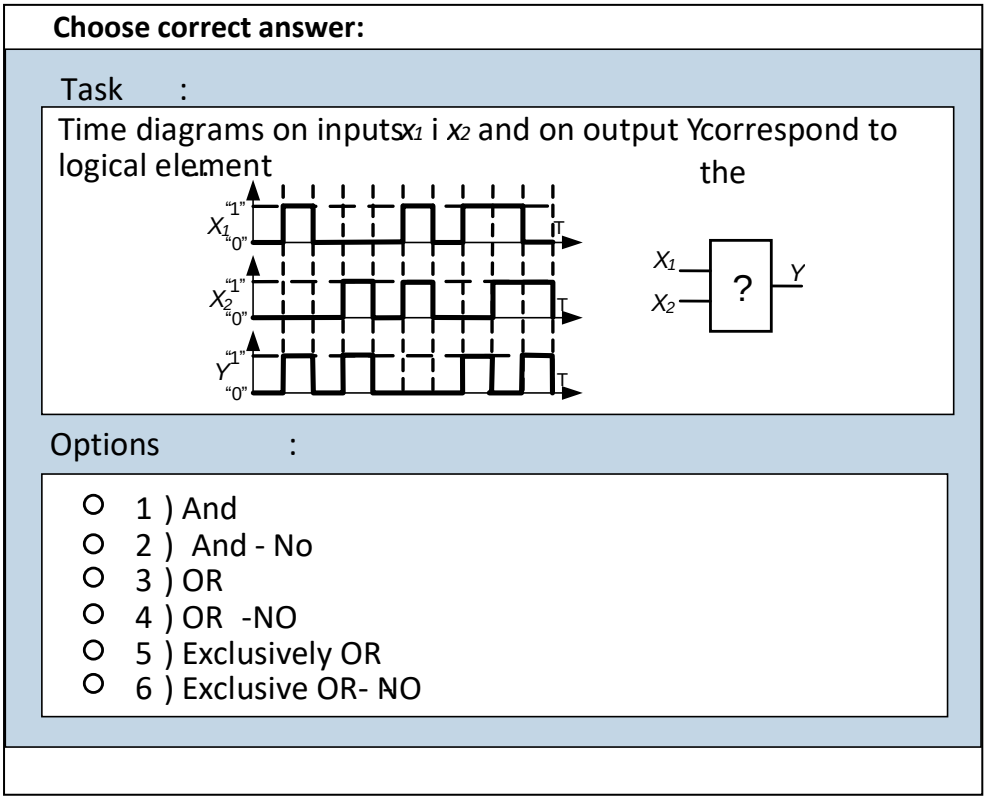

Figure 5. Facet test tasks with the choice of answers out of suggested options

Source: Own work

The form of the test questions with the choice of answer is a task for which there can be several options. The substantive basis of the tasks of this form is mainly classifying knowledge. The answers to the task must necessarily refer to the same kind or type of notions. In these test tasks the students should select all the correct answers from the list of options offered. In doing so, the instruction may indicate or not how many correct answers the student is giving. In the second case, the peculiarity of this form of task is that it is necessary to determine not only the correct answers, but also independently evaluate the completeness of the answer.

The principles of creating test tasks with multiple answer options are the same as the task of choosing one correct answer, only the number of proposed answers (distractors) can be increased at the discretion of the author to 5 - 12. But it is not necessary to make too cumbersome tasks as it is difficult for the student to keep too many content elements in memory at a time. Such a task can be regarded as combining n (by the number of answer options) of single-choice tasks, each of which independently of the other can be either true or false. The probability of complete guessing is equal to the product of the probability of guessing each question, which is p = 1/2. So:

$$p = \left(\frac{1}{2}\right)^n$$

For comparison: with n = 5, the probability of guessing is p = 3%, i.e. it turns out that with the same number of response options in the selection tasks, multiple choice provides a lower probability of random execution of the task than with the choice of one correct answer (p = 20%) .

The format of test tasks aimed at restoring the conformity of parts is a modification of multiple-choice test tasks and belongs to the category of logical pairs. Test tasks establish a measure of conformity of two heterogeneous large sets, which are in a known relation to each other. To establish correspondence between two objects (in this case - sets) - means to establish their dependence on each other; to establish a measure of conformity - means to identify this dependence in its entirety, that is, in terms of its comprehensiveness, completeness and uniqueness. Each set is formed by some set of elements, which can be any object, object, phenomenon, process, their components, properties and the like. The basic condition for the formation of the set is the homogeneity of all its elements, that is, the presence of common properties, signs.

A task of this type consists of a task statement and two or more columns. One column (left) is a list of initial conditions (words, phrases, sentences, dates, formulas, terms, etc.) to which the respondent should find the answer in the second column (right), which is called the list of answers (Figure 7). Response options are usually indicated by letters, and words or phrases containing the task are numbered.

The respondent should compare the material of the left and right columns and form the correct logical pairs. Consistently comparing each element of the set of initial conditions with the elements of the set of answers, the respondent finds out which of them are interrelated and which are not. This should be guided by the kind of relation between sets, which is determined by the condition of the problem and therefore, serves as a guide when finding interdependent elements.

The greatest difficulties in designing are related to the selection of plausible redundant elements in the right set. The plausibility of each distractor is empirically established. Each answer variant can be used more than once

or not used at all, so the task of this format cannot be executed by the exclusion method.

In fact, such test tasks are a task of choice; only the choice with respect to each of the n elements of the first set is made with the m variants of the answers contained in the second set. The probability of guessing:

$$p = \left(\frac{1}{m}\right)^n$$

For example: at n = m = 4, the guessing probability is p ≈0.4%

The assignments for establishing the correct sequence are used to test the knowledge of the sequence of certain actions, algorithms of execution, events in time, as well as definitions and concepts, etc. They help to form students' algorithmic thinking, knowledge and skills. Tasks of this form can be used as a means of controlling knowledge and skills, as well as a means of training (Figure 6).

At each step, one of the remaining correct answers is selected. The probability of complete guessing of all the answers of the task for N elements in the sequence will be equal to:

$$p = \frac{1}{n} \cdot \frac{1}{n-1} \cdot \ldots \cdot \frac{1}{2}$$

Thus, if the number of elements in the sequence is n = 5 the probability of guessing is p = 0.8%, and if n = 6; then p = 0.14%. It is possible to note the extremely low probability of guessing the correct answer characteristic of this form of tasks. However, in many cases the task of establishing the correct sequence is extremely non-technological or unsuitable because of the specific content of the subject matter. They are cumbersome and often allow for an ambiguous sequence of responses.

The variety of forms of test tasks used ensures their differentiation by the level of complexity, which, in turn, is necessary for differentiation of students by the level of academic achievement. A number of recommendations should be made:

1. An easy test task:

- is aimed at "learning" an object or to test "knowledge-recognizability";
- is aimed at choosing one answer from many with the knowledge of only one concept;
- the open type is aimed at revealing knowledge of the definition of a single-component base term;
- aimed at disclosing the basic concept;

– reveals the lowest level of the test specification hierarchy (for example, some "concept").

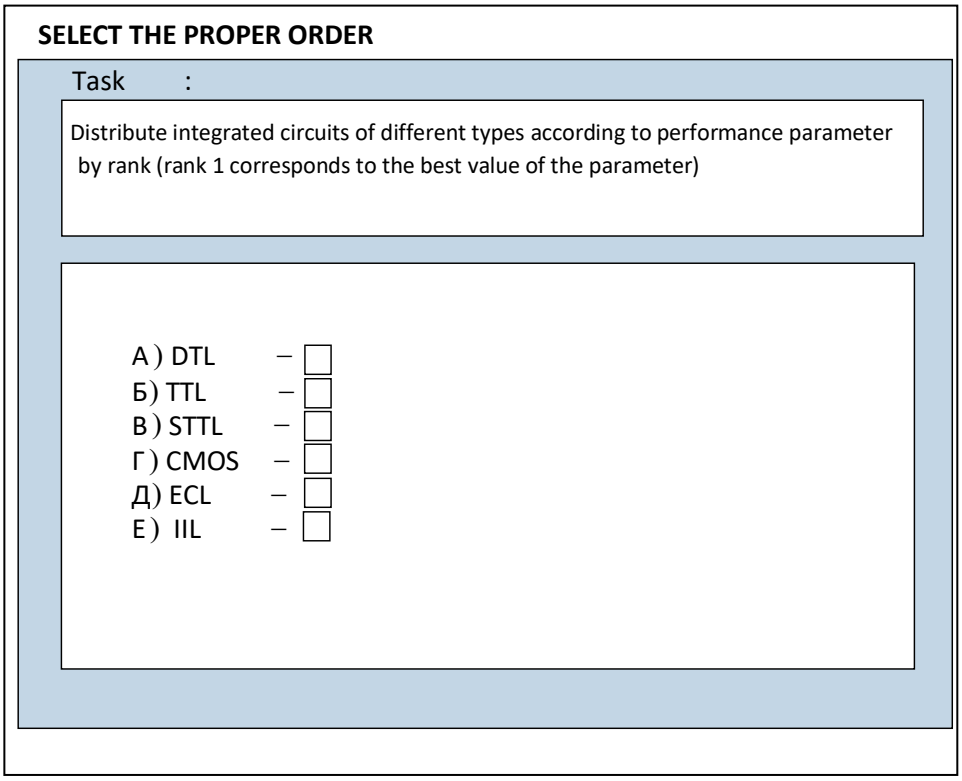

Figure 6. Test task to restore the order

Source: Own work

2. Tasks of medium level of complexity :

– are aimed at applying previously acquired knowledge in typical situations;

– reveals the average level of test specification hierarchy (for example, a topic or subtopic).

3. A complex level tasks:

– are aimed at applying the acquired knowledge and skills in non-standard conditions (i.e. in conditions not previously familiar to the subject);

– aimed at testing knowledge, skills and application;

– aimed at checking additional material;

– reveals the top level of the test specification hierarchy (for example, sections, chapters).

CONCLUSION

The offered method allows implementation of an individual approach in teaching, selection for the purpose of independent study exactly the educational material that is necessary for a particular student at a given point of time. The use of adaptive testing promotes the development of modern areas of education and opens up new opportunities to improve the efficiency of learning processes, provides more objective assessment of knowledge, skills and abilities, and also allows you to save time and, consequently, the cost of assessment. Implementation of this method should be balanced, so that this knowledge assessment procedure is well integrated into the learning process to ensure its maximum efficiency.

This testing technique requires considerable time to develop individual test sets, but the focus of the test on the content of individual classes allows you to implement this system for individual topics from the very beginning of its development. Further testing of the described methodology and statistical analysis of the obtained results is envisaged because only after conducting multiple statistical treatments can we talk about creating a test with stable quality parameters such as reliability and validity. The results of the research can be applied in the development of test sets for other courses.

REFERENCES

Abakumova, O.O., Bilec'kij, V.O. (2011). Sistema adaptivnogo testuvannya Student's Adapt. *Elektronika i svyaz' 4' Tematicheskij vypusk «Elektronika i nanotekhnologii»*, 196–202. [In Ukrainian]

Artamonov, Ye.B., Kashkevich, I.F. (2016). Metodika rozrobki adaptivnih elektronnih navchal'nih kursiv. *Problemi informatizacii ta upravlinnya,* 1 (53), 14–17. [In Ukrainian]

Babij, M.S., Chekalov, A.P. (2011). Primenenie elementov nechyotkoj logiki dlya rejtingovoj sistemy ocenki znanij. *Visnik Sums'kogo derzhavnogo universitetu. Seriya: Tekhnichni nauki,* 3, 116–121. [In Ukrainian]

Conejo R., Guzmán E., Millán E., Trella M., Pérez-De-La-Cruz J.L., Ríos A. (2004). Siette: A web-based tool for adaptive testing. *International Journal of Artificial Intelligence in Education,* 14 (1), 29-61.

Cidelko, V.D., Yaremchuk, N.A., Shvedova, V.V. (2009). Avtomatizovana sistema testuvannya, navchannya ta monitoringu. Pat. 43616 Ukraïna: MPK G09B 7/00. Zamovnik ta patentovlasnik: Nacional'nij tekhnichnij universitet Ukraïni «Kiïvs'kij politekhnichnij institut». № 200902620, zayavl. 23.03.2009, opubl. 25.08.2009, Byul №16. [In Ukrainian]

Fedoruk, P.I. (2007). Adaptivni testi: statistichni metodi analizu rezul'tativ testovogo kontrolyu znan'. *Matematichni mashini i sistemi*, 3/4, 122–138. [In Ukrainian]

Katerinchuk, I. S., Kravchuk, V. V., Kulik, V. M., Rachok, R. V. (2009). Intelektual'na sistema avtomatizovanogo kontrolyu znan' studentiv vishchih navchal'nih zakladiv. *Informacijni tekhnologiï v osviti,* 4, Herson, Vid-vo HDU, 139 –147. [In Ukrainian]

Kovaleva, Ye. A. (2014). Razrabotka kompleksa interaktivnyh testov po matematike v Adobe Captivate. *Tekhnologicheskij audit i rezervy proizvodstva,* 6(3), 51–55. [In Russian]

Kravchenko, O.M., Plakasova, Zh.M. (2010). Model' intelektual'noï kontrolyuyuchoï pidsistemi z bagatorivnevim adaptivnim testuvannyam. *Skhidno-Єvropejs'kij zhurnal peredovih tekhnologij,* 4/2' (46), 21–25. [In Ukrainian]

Kravchenko, O.V., Plakasova, Zh.M. (2010). Aspekti formuvannya testiv dlya kontrolyu znan' v sistemi adaptivnogo navchannya. *Shtuchnij intelekt,* 4, 576–583. [In Ukrainian]

Louhab F.E., Bahnasse A., Talea M. (2017). Towards a contextual mobile learning deployment: An overview. *International Journal of Computer Science and Network Security,* 17 (7), 80-88.

Lizunov, P.P., Teslya, Yu.M., et al. (2009). Organizacijni aspekti sistemi testuvannya dlya promizhnogo i pidsumkovogo kontrolyu znan' studentiv. *Inform. tekhnologiï v osviti:* zb. nauk. pr., 4., 124–133. [In Ukrainian]

Nacional'na doktrina rozvitku osviti Ukraïni u XXI stolitti: Proekt. Osvita, 2001, 60-62, 24 – 31 October.

Naumenko, S. (2016). Zarubizhnij dosvid zastosuvannya komp'yuternogo adaptivnogo testuvannya. *Pedagogichna komparativistika – 2016*: osvitni reformi ta innovaciï u globalizovanomu sviti: materiali nauk.-prakt. seminaru (Kiïv, 06 chervnya 2016), In-t pedagogiki NAPN Ukraïni, za zag. red. O. I. Lokshinoï, Kyiv, Pedagogichna dumka, 192–196. [In Ukrainian]

Popov, D.I., Lazareva, O.Yu. (2015). Nechetkaya overlejnaya model' uchashchegosya v intellektual'noj obuchayushchej sisteme. *Nauchnyj vestnik MGTU GA.;* (213), 141–148. [In Russian]

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. With a Foreword and Afteword by B.D. Wright. The Univ. of Chicago Press. Chicago & London, 199 p.

Rizun, N. O., Taranenko, Yu. K. (2013). Poluchenie matematicheskoj modeli intellektual'noj deyatel'nosti testiruemogo s primeneniem statisticheskih metodov. *Sistemni tekhnologiï*, 3, 97–108. [In Russian]

Rizun, N.O. (2012). Model' pidvishchennya motivaciï studentiv do navchannya (aspekti organizaciï ta provedennya testovogo seansu). *Visnik NTU "HPI". Tematichnij vipusk: Informatika i modelyuvannya,* Harkiv: NTU "HPI", 38, 148–155. [In Ukrainian]

Shcherbina, D.N. (2015). Strategii prohozhdeniya testov znanij, vyyavlennye metodom hronometrii prosmotra variantov otveta. *Valeologiya*, 4, 112–121. [In Russian]

Tseng W.-Ta (2016). Measuring English vocabulary size via computerized adaptive testing. *Computers & Education*, 97, 69-85.

Veligura, A.V., Lekhcier, L.R., Tkachenko, V.P. (2003). Sposib vimiru rivnya znan' uchniv pri komp'yuternomu testuvanni. Pat. 61415 Ukraïna: MPK 7G06F7/00 / Zamovnik ta patentovlasnik: Skhidnoukraïns'kij nacional'nij universitet imeni Volodimira Dalya. № 003010849, zayavl. 31.01.2003, opubl. 17. 2003. Byul № 11. [In Ukrainian]

Zagrebel'nij, S., Zagrebel'na, O., Kostikov, O. (2016). Metodi adaptivnogo testuvannya znan' studentiv. *Pedagogichni nauki: teoriya, istoriya, innovacijni tekhnologiï,* 3, 376–384. [In Ukrainian]

Zyanchurina, I.N., Nekrasov, V.I. (2011). Organizaciya testirovaniya s uchetom stepeni slozhnosti testov. *Otkrytye informacionnye i komp'yuternye integrirovannye tekhnologii:* sb. nauch. tr., Kharkiv, Nac. aerokosm. un-t "HAI", 50, 111–117. [In Russian]