

## Construction of Expert Knowledge Monitoring and Assessment System Based on Integral Method of Knowledge Evaluation

Viktoriya N. Golovachyova<sup>a</sup>, Gulbakhyt Zh. Menlibekova<sup>b</sup>, Nella F. Abayeva<sup>a</sup>, Tatyana L. Ten<sup>c</sup>, Galina D. Kogaya<sup>a</sup>

<sup>a</sup>Karaganda State Technical University, Karaganda, KAZAKHSTAN;  
<sup>b</sup>L.N. Gumilyov Eurasian National University, Astana, KAZAKHSTAN;  
<sup>c</sup>Karaganda Economic University, Karaganda, KAZAKHSTAN

### ABSTRACT

Using computer-based monitoring systems that rely on tests could be the most effective way of knowledge evaluation. The problem of objective knowledge assessment by means of testing takes on a new dimension in the context of new paradigms in education. The analysis of the existing test methods enabled us to conclude that tests with selected response and expandable selected response do not always allow for evaluating students' knowledge objectively and this undercuts the effect of pedagogical evaluation of their cognitive activity as well as the teaching and learning processes generally. Authors propose an expert knowledge monitoring and evaluation system based on an integral method of knowledge evaluation. This method is built on a new approach to constructing test items and responses to them, which give students an opportunity to freely construct their responses, and presupposes a set of criteria for their assessment. Proposed method makes it possible to expand the functions of tests and in this way approximate the test grade to the real level of students' knowledge. Theoretical and empirical data presented in the paper can be used for improving the monitoring and evaluation of knowledge in social sciences and the humanities and thus raising the quality of education.

### KEYWORDS

Quality of education, expert knowledge monitoring and evaluation system, integral method of knowledge evaluation, forms of constructing a test response

### ARTICLE HISTORY

Received 20 April 2016  
Revised 28 April 2016  
Accepted 9 May 2016

### Introduction

Modern information technologies have long been part of higher education process. However, many conceptual issues of the design of teaching and testing software systems have not been resolved yet (Zaytseva et al., 2013).

**CORRESPONDENCE** Viktoriya N. Golovachyova ✉ golovacheva\_vn@mail.ru

© 2016 Golovachyova et al. Open Access terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>) apply. The license permits unrestricted use, distribution, and reproduction in any medium, on the condition that users give exact credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if they made any changes.



IT progress in all walks of life, including education, shows that computer-based expert systems are becoming an important part of education.

Designing academic tests is a considerable part of the design of knowledge monitoring and evaluation systems. This method allows for interpreting results of training with a high degree of objectivity, while being an effective, rational and convenient tool of knowledge assessment.

Scientists agree that modern research-based didactics should be based on a rich battery of maximally objective methods of pedagogical diagnostics. However, we support the opinion of V.V. Burlai, N.E. Suflyayeva & L.R. Yurenkova (2012), that “computer-based assessment is reasonable in the epoch of information technologies... It allows for meeting a very important requirement, i.e. of uniform standards in knowledge assessment”.

Modern systems use classical forms of testing, namely items with selected responses or expandable selected responses. This does not always allow to assess the real level of knowledge. In these cases, there is no possibility for students to freely construct their responses, which would provide for a more accurate evaluation (Golovacheva & Abaeva, 2015).

International practice shows that the existent methods of constructing test items and responses do not always allow to assess students' knowledge. This is especially true for social sciences and the humanities, where tests provide for gross distortions and stereotyping of material as well as rote learning by students.

The current level of IT allows for constructing tests with multiple-choice, true-false, numerical, constructed or other responses. Standardized tests are the most widespread. They are easier to create as there is no necessity to foresee every possible correct answer. The main advantage of these tests is that they are easy to use. With standardized tests students spend their time and effort on the task rather than on putting their answers down on paper.

Besides, as a group of scientists affirm, it is necessary to take into account content validity, the logical structure and forms of the test items, which allow for computer processing; the quality of tests according to given parameters (Atoev, Valisheva & Khamidov, 2015; Yarullin, Prichinin & Sharipova, 2016).

As years of tests' use in the educational process show, this form of assessment has many advantages. Among them is the scope of student coverage, the simplicity and efficiency of grading; the possibility of complete computerization of the testing procedure; decrease in the subjectivity of assessment.

In order for knowledge assessment via testing to be successful, it is necessary to keep track of knowledge acquisition at every stage of learning (Talbi, Warin & Kolski, 2013; Kamalova & Raykova, 2016). At the same time, it is necessary for a test to embrace all the characteristics of knowledge acquisition. Tests should embrace such parameters as knowledge of facts, an ability to illustrate one's answer with examples, the skill of cohesively and concisely expressing one's ideas etc. Only such techniques of testing that are not inferior to oral examination allow to take full advantage of a test (Tsaritsentseva, 2013).

In addition, one important problem is that a computer-based test cannot fully grasp the complexity of a textual answer. Conventional methods of analysis of computer systems are based on precision machining of numerical data, and are not capable of comprising the enormous complexity of human thinking and decision-making.

It can be asserted that in order to assess progress in such a humanist system as education, it is necessary to overlook the high standards of accuracy and rigor that we expect, for instance, from mathematical analysis of clearly defined mechanical systems, and be tolerant of other options of assessing knowledge.

The accuracy and objectivity of knowledge evaluation depends not only on the construction of test responses, but also on the criteria at its core, the parameters of students' knowledge evaluation and the grading scale and the rating system used (Liu, Lan & Ho, 2014).

It is possible to achieve substantial results only by using such techniques of testing that are not inferior to an oral examination/interview. Objectification of knowledge assessment can be achieved in this case only given a set of criteria for knowledge assessment and the rules of their determination.

In spite of the progress in this sphere, the analysis of expert knowledge assessment systems has unearthed certain problems in their use. The problem of objective knowledge evaluation is accounted for by the multidimensionality of this issue from the point of view of teachers, psychologists and methodologists (Golovacheva & Abaeva, 2015).

We maintain that such expert systems have opportunities for more efficient knowledge evaluation, but currently the relevant algorithms and their programs are not implemented. One of the reasons is that such computer systems have difficulty identifying freely constructed responses. Therefore most expert systems use the universal and simple method of analysis based on key words, which compares the answers with previously obtained samples.

Comparison against a sample does not necessarily require the response's full coincidence with the sample. There are possibilities for finding out partial coincidence or coincidence with one sample out of the whole set. This considerably extends the system's opportunities as far as processing the test-takers' responses is concerned and makes it more "intelligent" as a whole.

Clearly, innovation is always linked to risks since it is impossible to always predict the ultimate outcome and avoid false assumptions. Innovation should be carefully elaborated, designed and organized (Fidalgo-Blanco, Sein-Echaluce & García-Peñalvo, 2014).

Improvement of the method of testing is possible by means of a new approach to constructing responses to test items, i.e. permitting a free form of response. Identification of these responses can be done using a database elaborated by experts in a relevant field of knowledge. The responses' evaluation should be carried out by applying specific criteria based on the algorithms for evaluating the quality of these responses.

The problem of objective assessment is one of the pressing issues in the theory and practice of education. Despite the progress made in this field, the issue of adequate appraisal of students' performance by means of grades is still open.

The object set by this paper is search for pedagogically effective ways and elaboration of a method that permits to improve the process of knowledge assessment with a view to raising the quality of education.



## Literature Review

Tests are a popular method of knowledge assessment and designing computer-based tests in particular has become a form of art. They offer an opportunity to standardize knowledge assessment and exclude the negative implications of human factor, save time of both students and examiners (Tsai et al., 2015).

A system of concepts and terms as well as form and content makes up the theoretical and methodological bases for test design. Computer-based tests are a convenient tool for knowledge assessment, particularly in the educational process. Testing is an important element of control of this process (Kim & Jang, 2014; Morris & Chikwa, 2014).

According to V.S. Avanesov (2012), tests have a doubtless advantage over other methods of knowledge assessment. This scholar, who is a test expert, emphasizes five main advantages:

1. High scientific validity of testing, which allows to receive the objective evaluation of the test-takers' level of knowledge.
2. Technological effectiveness of test methods.
3. Accuracy of measurement.
4. Uniformity of the rules of testing and interpretation of test results for all users.
5. Compatibility of the test technology with other modern education technologies.

Tests are not a new method of knowledge evaluation. Psychologists started using tests to learn about individual differences of people already in the 1880s (F. Galton, D. Kettel). F. Galton was one of the creators of the scientific method of tests. He made a considerable contribution to the theory of tests and extended their practical application based on mathematic-statistical methods, shaped by him in metrology. He also proposed the concept of correlation, which is still applied in science (Kadnevskiy, 2012).

At the same time, tests are the least theoretically and practically elaborated form of assessment today, as the analysis of scientific-pedagogical literature and educational practice show (Zaytseva, Smorodina & Vasina, 2013; Ibragimov et al., 2016).

The advantage of this form of assessment lies in its objectivity, i.e. the independence of evaluation from the expert (teacher) possessing knowledge of a given subject area (Waight, Chiu & Whitford, 2014). However, it can be argued that tests do not always provide an objective appraisal of a knowledge level. There arises a question: how did a test-taker come up with the correct answer: by means of logical reasoning or accidentally? Besides, there is always a chance that a student has learnt the material by rote.

Experience shows that rationally composed tests require a response in one of the following forms:

- 1) selection of the correct answer from a series of options which are right/wrong, complete/incomplete, accurate/inaccurate;
- 2) selection in two parts (which presupposes making a choice in the first part of a test and explaining it in the second part);
- 3) selection of one of two options (yes/no, 0/1, true/false);

- 4) putting suggested components in the right order;
- 5) matching (two lists of components);
- 6) completion of a sentence;
- 7) one-word/number answer;
- 8) multi-word answer with restrictions as to the order or connections between words.

The expedience of tests is beyond doubt. They are certainly the most advanced and one of the most popular forms of knowledge assessment today (Martinez-Garza & Clark, 2013).

At the same time, it is necessary to point out the negative sides of testing as well. It does not always allow for gauging the knowledge level of students objectively, especially in social sciences and the humanities (Ji, 2013). It hampers students' perception of subject matter as cohesive, provides for stereotyping of knowledge, lack of creativity, rote learning. Tests allow for cheating in case students have the test keys, and the tests' reliability rates depend on the variability of test scores in different groups of students etc. (Bloemeke, Koenig & Busse, 2014). Moreover, test-takers cannot answer as they wish, which is an important criterion in objective knowledge evaluation (Golovacheva, Abaeva & Kokkoz, 2015).

According to Sorokina and Kolobova (2014), tests have a number of drawbacks. The principal one is the impossibility of checking students' speech culture (written and oral), narrowness of the subject content in students' minds.

This has obvious negative consequences: decrease in the stimulating effect of knowledge assessment on students' cognitive activity and the quality of the educational process in general.

The slow rate of elaboration of new knowledge assessment methods is the main reason for the gap between the current and potential capacities of expert knowledge assessment systems.

### **Aim of the Study**

The main aim of this research is to define ways of teaching effectiveness improvement and as well as to develop a method aimed at knowledge monitoring and assessment to make the quality of education better.

### **Research questions**

The overarching research questions of this study was as follows:

What is a new integral method of knowledge monitoring and evaluation system?

What is the design of test questions and answers in academic tests?

What are the benefits of developed expert system of knowledge assessment and control based on Integral Method of Knowledge Evaluation (IMKE)?

### **Method**

This research is based on the dialectic method of scientific cognition and a systems approach. In the course of the investigation we applied such general scientific methods and techniques as scientific abstraction, analysis, synthesis, prediction and others.



The research was carried out with the application of *theoretical methods*, which were necessary for defining the problem and analyzing the collected data. The study relies on the works of methodologists on the issues of knowledge assessment based on tests and of pedagogics in particular, works on general and specific pedagogics; *on traditional pedagogical methods*: inductive and deductive methods, study and theoretical analysis of scientific, psychological and pedagogical and methodological literature on the artificial intelligence systems; analysis and generalization of advanced pedagogical experience, pedagogical observation, conversations, questionnaires, experiments with students and teachers; *diagnostic methods*: observation, survey design, interview, knowledge assessment by tests; *scientific methods*: analysis, synthesis, generalization.

*Testing* helped to draw a statistical portrait of the changes in the academic progress of students and identify their achievements, to see how this or that form of knowledge assessment affected them. It also helped to express qualitative changes in numbers, which contributed to better understanding of the efficiency of research methods we applied.

*Mathematical and statistical methods* were applied for processing the data obtained by means of questionnaires and experiments and for identifying quantitative relations between the studied phenomena. They helped to evaluate the experiment results, enhance the reliability of conclusions and gave grounds for theoretical generalizations.

The following methodologies were widely applied within the statistical method: registration, rating, scaling and nominal scales.

*Pedagogical prediction*, linked to the definition of objectives, was used with the purpose of specifying pedagogical aims and their transformation into a system of set scientific-pedagogical tasks.

*Simulation* (modelling) is a more powerful transformational means of pedagogical research. A scientific model is a visualized or materialized system which represents the subject of research and is capable of substituting it in such a way that the study of this model provides new information on the object. The main advantage of this method is the integrity of the information presented. Simulation is based on synthesis, i.e. isolating whole systems and researching their functioning.

By applying the simulation method we achieved the following three goals: heuristic – for the classification, designation, discovery of new laws, development of new theories and interpretation of the obtained data; computational – for solving computational problems by means of models; experimental – for solving the problem of empirical testing (verification) of the hypothesis by means of dealing with this or that model.

The experimental part of the work is based on numerous methods of computer simulation and simulation experiments with the use of high-level programming language. We also employed mathematical statistics methods while processing the experimental research results.

The authenticity of the ideas, conclusions and practical recommendations present in the study is confirmed by completeness and rightness of the original assumptions, a theoretical substantiation based on the use of a rigorous mathematical apparatus, virtually complete coincidence of the theoretical results with the results of the experiments and implementation of the obtained results.

## Data, Analysis, and Results

Expert knowledge assessment system *IMKE* is a computer -assisted program based on the integral estimation of knowledge.

A method in which objectification of knowledge assessment is achieved using a set of criteria of its formation is described as integral. The essence of such a method is that students' knowledge and skills are evaluated on the basis on test items to which they respond in a freely constructed form.

A schematic diagram of expert knowledge assessment system *IMKE* is presented in Figure 1.

The core of the system is the basic vocabulary of the subject area, which is created and accumulated in the process of its formation and teaching and is the main didactic reference material for the analysis of a test response.

The input data of the tested person are fixed and registered in a lexical analyzer of *IMKE*. The lexical analyzer receives the source answer text directly from elements of the input interface and converts it into an array of lexical units (a word or a number). White spaces, hyphens, line end characters etc. are deleted. The analyzer carries out a search procedure for every word of the response using the basic vocabulary of the subject area. In case of an exact match of the analyzed word with a word from the vocabulary, the search is considered to be successful and the corresponding information is transferred to the program. On the basis of this information, a set of criteria are established that characterize the quality of a student's acquisition of knowledge, namely subject matter  $\delta$ ; literacy  $\gamma$ ; provision of examples  $\varphi$ ; coherence  $\mu$ ; complexity  $\eta$ . (Figure 1).

The algorithm of establishing each of the selected criteria is as follows.

### **Subject matter**

The criterion reflects the basic level of subject knowledge and is determined by a correspondence of the used terms with the thesaurus. It is a component part of the subject area's basic vocabulary. The evaluation criterion is the ratio of the number of properly used key words to the total number of key words corresponding to every test item in the thesaurus.

The basic vocabulary of the subject area is the main didactic reference material, which serves for the analysis of every test response.

The computer procedures of the establishment of criterion "subject matter" are the following:

- a. The text of the test response is broken up into words.
- b. Every word of the response is checked for correspondence to thesaurus key words and their synonyms.
- c. The amount of corresponding key words is calculated.
- d. The criterion calculation:

$$\delta = N / M$$

where N is the quantity of key words corresponding to the thesaurus as regards every test item;

M is the total amount of key words corresponding to the thesaurus as regards every test item.



**Literacy** is determined by grammar rules of constructing text documents. The literacy of an answer is established according to the following parameters:  $\gamma_1$  – presence of subject;  $\gamma_2$  – presence of predicate;  $\gamma_3$  – presence of object(s);  $\gamma_4$  – presence of attributive(s);  $\gamma_5$  – presence of adverbial modifier(s);  $\gamma_6$  – absence of grammar mistakes;  $\gamma_7$  – absence of errors of style;  $\gamma_8$  – presence of diverse syntactic constructions.

The selected coefficients characterize the answer as a composite literacy index, thus, for example: A – complete sentence;  $\gamma_1 \cap \gamma_2 \cap \gamma_3 \dots \cap \gamma_q$ , where  $q$  is the quantity of structural-semantic components determining this index.

If  $q = \max I$ , where  $I$  is the maximum quantity of structural-semantic components for which the condition is fulfilled, then the answer is considered to be complete.

B – incomplete sentence;

$\gamma_1 \cap \gamma_2 \cap \gamma_3 \cap \gamma_q$ , where  $Q < I$ ;

C – limited sentence: only  $\gamma_1 \cap \gamma_2$ ;

D – incorrect sentence: absence of either  $\gamma_1$ , or  $\gamma_2$ ;

E – ungrammatical sentence: grammar mistakes;

F – incorrectly shaped sentence: presence of errors of style.

A set of rules determining the general grammar quality of a textual response can be an evaluation criterion.

The algorithm of determining the criterion of literacy is the following:

1. The text of the response to the test item is broken up into words.
2. The spelling of every word is checked against a dictionary.
3. The quantity of correctly written words in the response text is calculated.
4. The total amount of words in the response text is determined.
5. The criterion calculation.

$$D = Q / R$$

where  $Q$  is the quantity of correctly written words in the response text;

$R$  is the total amount of words in the response text.

### **Provision of examples**

Examples illustrate answers to the questions posed in a test. The criterion is determined by establishing the correspondence of the examples used in the response text to the words in the example database or their synonyms. The criterion of evaluation is the ratio of the quantity of correctly provided examples to the total amount of examples corresponding to their database on every test item.

Depending on the teacher's experience, evaluation criteria may include: provision of one example, provision of two or more examples, absence of example, incorrect example.

The algorithm for determining the criterion "provision of examples" is the following:

1. The response text is broken up into words.

2. Every word is checked for correspondence to the database of examples or their synonyms.
3. The quantity of example words corresponding to the example database is calculated.
4. The criterion calculation.

$$F = D / F$$

where D is the quantity of examples corresponding to the example database or their synonyms according to every test item;

F is the total amount of examples from the example database and their synonyms according to every test items.

### **Sentence coherence**

The text of the response to the main question in the test must consist of a few sentences making a meaningful whole. There must be a gradual idea development in the sentences. The first sentence is usually described as a generalizing one. The remaining sentences must cohere with the main sentence, or all the sentences except the main one must be subordinate to the generalizing sentence, i.e. complete its meaning, reveal and expand on the essence of the generalizing sentence, classify the subjects of the question etc.

The algorithm of determining the criterion of sentence coherence is the following:

1. The text of response to the main test item is broken up into sentences.
2. The quantity of sentences in the response text is calculated.
3. The response incidence matrix is filled according to the principle shown in Figure 2.

*Note:* n – quantity of sentences in the response text; m – quantity of key words corresponding to the thesaurus according to the test item or their synonyms; X – parameter allowing to establish sentence coherence.

4. The maximum possible quantity of links between the sentences used in the response text is calculated according to the following formula:

$$L = n \cdot (n - 1) - \sum_{i=1}^{n-1} i$$

where n is the quantity of sentences in the response; parameter i = 1, 2,..., n-1.

The criterion calculation.

$$M = E / L$$

where E is quantity of links between sentences in the response;

L is the maximum quantity of links between sentences in the response.

### **Complexity**

This criterion applies to the quality of every test response in general and is established based on the presence of links between the specified criteria: subject matter, literacy, provision of examples, sentence coherence. If a test-taker, while responding to the main test item, puts the criterion “subject matter” quite highly, provides examples, explaining the essence of the question, and his/her answer text

is grammatically correct, and there is a gradual idea development in the answer, then this answer is complex from the point of view of general characteristics.

The analysis of complexity is carried out on the basis of quantity of connections in a response and is determined in an expert way. Every test item is graded according to the following formula:

$$\text{Ball} = [(\bar{\delta} + \bar{\gamma} + \bar{\varphi} + \bar{\mu} + \bar{\eta}) / w \cdot \kappa]$$

where  $\bar{\delta}$  – grade for “subject matter”;

$\bar{\gamma}$  – grade for “literacy”;

$\bar{\varphi}$  – grade for “provision of examples”;

$\bar{\mu}$  – grade for “sentence cohesion”;

$\bar{\eta}$  – grade for “complexity”;

w – total amount of the criteria of answer analysis ;

$\kappa$  – complexity index of every test item.

The accuracy and objectivity of knowledge assessment depends not only on the construction of test responses but, among other things, on the criteria at its core, the parameters that are identified for assessing students’ knowledge and the grading scale or rating system used.

Every selected criteria is graded according to a ten-point scale. The total grade is calculated taking into account the statistical grade structure obtained by the *IMKE* analyzer.

The key problem in providing the final knowledge assessment is determining the borders between two grades, when knowledge can be evaluated somewhat higher or lower than a certain grade. Fuzzy sets for criterion scores were used in the study to establish the borders of grade evaluation, and the heuristic method was applied in determining the final grade. While using fuzzy sets, a preference function is composed, confidence coefficients are selected and borders of the fuzzy sets are determined for every grade before determining the grade according to a five-point system.

Within the framework of the heuristic approach to knowledge assessment, we elaborated an algorithm, which is a problem in combinatory analysis. It is defined as follows: it is necessary to determine all the possible values  $\alpha$ ,  $\beta$  and  $\lambda$  in such a way that they meet the conditions of getting a grade on a five-point scale:

$$a \leq \alpha_k + \beta_m + \lambda_n \leq d \text{ (excellent)}$$

$$B \leq \alpha_k + \beta_m + \lambda_n < a \text{ (good)}$$

$$c \leq \alpha_k + \beta_m + \lambda_n < B \text{ (satisfactory)}$$

$$0 \leq \alpha_k + \beta_m + \lambda_n < c \text{ (unsatisfactory)}$$

with extreme values k, m, n;

where  $\alpha$ ,  $\beta$ ,  $\lambda$  are all possible combinations of complicated, average and simple questions respectively: k is the total amount of complicated questions in the test; m is the total amount of average questions in the test; n is the total amount of simple questions in the test; d is the total amount of questions in the test; a, b, c are amounts of the correct test answers fulfilling the conditions for getting the grade according to a five-point system.

### **Knowledge assessment system IMKE in practice**

We have checked the accuracy of knowledge assessment results using expert knowledge evaluation system *IMKE*. A hypothetical “perfect” knowledge evaluation system was chosen as a benchmark and was called theoretical. After comparing the knowledge assessment results obtained by testing (control group) and the results obtained by the integral method of knowledge evaluation (experimental group) with the results of a theoretical evaluation, we observed greater accuracy and efficiency in the work of *IMKE* expert evaluation system. The basis of this system is the integral method of knowledge evaluation proposed by us. It was assumed that the closer the parameters of the expert system under investigation are to those of the theoretical system, the more advanced it is.

In striving for the accuracy of results we wanted to establish: the interrelation between the considered methods of knowledge evaluation (test method (T) and integral method of knowledge evaluation (I)); statistical characteristics of evaluation results according to types of assessment: average grade (T<sub>aver</sub>; I<sub>aver</sub>), standard deviation (σT; σI); whether the sample data correspond to the hypothesis of probability distribution of general totality (by applying K. Pearson’s chi-squared test at the level of significance 0.05).

The obtained results of the statistical processing of experimental data show that the functions of the grades distribution in the control and experimental groups are close and obey the same law. Nevertheless, the distribution function of grades obtained in the control and experimental groups, is closer to and obeys the same law as the distribution in the theoretical group (the theoretical frequency is less than the critical values obtained in the processing of data in the control groups ( $P_t(\chi^2 \geq \chi_q^2) = 0.0047$ ; at  $k = 1$  and  $\chi_q^2 = 7.514485$ ;  $0.0047 < 0.005$ ;  $P_i(\chi^2 \geq \chi_q^2) = 0.0833$  at  $k = 1$  and  $\chi_q^2 = 2.985654$ ;  $0.0833 > 0.005$ ).

Thus, we have achieved an increase in the quality of education by means of the elaborated expert knowledge assessment system *IMKE*, the basis of which is the integral method of knowledge evaluation. The increase in the quality of education was due to the obtaining of objective information by the teacher about the level of knowledge acquisition by students; detailed analysis of the content of knowledge assessment, which increases the interest in and motivation to pursue education. This is confirmed by the data that are available for the teacher after the test. We have also achieved task-oriented correction of the education process taking into account test results, selection of complexity factors, changing the borders of the unclear grade definition while determining the final grade.

### **Discussion and Conclusion**

Expert computer-assisted knowledge assessment systems, based on didactic tests and on various approaches to grades’ assignment and aimed at providing high-quality education, are becoming increasingly popular.

In this regard, the tasks related to the criteria of assessing educational activity are some of the most challenging ones in modern pedagogics.

The analysis carried out as part of this study suggests that tests items requiring responses of the selected and selected-expandable type do not always provide an opportunity of evaluating students’ knowledge objectively, especially in social sciences and the humanities. This situation has obvious negative consequences, for instance, a decrease in the stimulating effect of knowledge testing



on the students' cognitive activity and the educational process in general. Of special relevance is the need of creating an expert knowledge assessment system which allows to reveal students' real level of knowledge in social sciences and the humanities, i.e. the subjects, where emphasis is placed on human knowledge and reflection.

The pace of addressing methodological problems and creating new knowledge assessment methods falls short of the opportunities of expert computer-based knowledge assessment systems. Didactics of the 21<sup>st</sup> century strives for control and appraisal of the educational process at every stage, from the elaboration of aims and content to checking results. That is the reason for continuous intensive search of ways and means of improving knowledge assessment with a view to enhancing the quality of education (Elliot, Wilson & Boyle, 2014).

The results of this study can be used in general pedagogics as well as in theoretical and practical testology. The paper substantiates the need to take a new approach to responding to a test item, i.e. a freely constructed test item and test response, as well as the necessity for elaborating the criteria of analyzing such responses and for a research-based approach to their evaluation. The processing of the entire information contained in test responses is carried out with the help of algorithms of analyzing test responses and computer means of processing data. This offers an opportunity to get an all-round objective evaluation of knowledge. The test procedure is rigorously formal, but its result proceed from the responses given by the test-takers.

The practical significance of the research consists in the fact that we have set up an expert system of knowledge assessment, *IMKE*, which can be used for improving knowledge evaluation in social sciences and the humanities and enhance the quality of education in general. It provides for solving the scientific problem of objective and accurate knowledge assessment by means of an expert computer-based system of testing.

Many years of research by various authors suggest that a grade that represents the level of a group student's knowledge must be normally distributed. Therefore the most effective system of knowledge assessment is the one that does not overstate or skew the average grade in a group's responses. This implies that the hypothesis of the normal distribution of grades in the monitoring of the education process is the main working hypothesis (Van den Hurk et al., 2014).

Using this hypothesis in our work, we checked the veracity of the results of knowledge assessment by means of expert system of knowledge assessment *IMKE*.

The results obtained in the study do not address all the aspects of the problem of quality of knowledge obtained in the process of education. Further theoretical and practical elaboration of this subject requires solution of such problems as improving the integral method of assessment as regards an increase in the quantity of criteria of knowledge assessment, elaboration of criteria of assessing them, development of a knowledge base, involvement of various kinds of analyzers etc.

### Implications and Recommendations

The analysis showed that the selected design of test questions and answers (questions that implied answers of selective and selective-constructed types) do not always provide objective assessment of the students' knowledge. This situation has obvious negative consequences: reduced stimulating effect of assessment on cognitive activity of students, as well as on the quality of the entire training process.

What seems especially important - the need to establish an expert system of knowledge assessment and control that would determine the actual level of student knowledge related to the social and humanitarian subjects. The expert system of knowledge assessment and control IMKE provided effective solution of these problems.

The originally developed expert system of knowledge assessment and control IMKE was put into the learning practice; this system can be recommended to improve knowledge assessment and control as regards social and humanitarian subjects with a view to improve the training quality. This enables using the research results to solve the scientific problem of objective and reliable knowledge assessment by using expert system of knowledge assessment and control.

The paper theoretically justified the need for a new approach to finding answers to the test question, allowing free - constructible form of test questions and answers, as well as the need to develop the result analysis criteria and the scientifically based approach to their assessment. Processing of full test result data is carried out through the developed algorithms for calculating the criteria for test result analysis, and software tools providing a comprehensive and objective assessment of knowledge. The pedagogical testing procedure is strictly formalized in this regard; however, the results become clear from the student responses.

The developed expert system of knowledge assessment and control IMKE, based on the integral method of knowledge assessment, provided the improved training quality through obtaining the objective information on the degree of knowledge assimilation by students. The interest and learning motivation of students were significantly increased

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Notes on contributors

**Viktoriya N. Golovachyova** holds a PhD in Pedagogy and now is an Associate Professor at Karaganda State Technical University, Karaganda, Kazakhstan.

**Gulbakhyt Zh. Menlibekova** holds a PhD and now works at L.N.Gumilyov Eurasian National University, Astana, Kazakhstan.

**Nella F. Abayeva** holds a PhD in Pedagogy and now is a Senior Lecturer at Karaganda State Technical University, Karaganda, Kazakhstan.

**Tatyana L. Ten** holds a PhD in Technical Science and now is a Professor at Karaganda Economic University, Karaganda, Kazakhstan.

**Galina D. Kogay** holds a Ph.D and now is an Associate Professor at Karaganda State Technical University, Karaganda, Kazakhstan.

### References

- Atoev, E. Kh., Valisheva, N. A. & Khamidov, E. E. (2015). Quality of Testing as a Basis for Objective Assessment of Students' Knowledge. *Young Scientist*, 3, 725–27.
- Avanesov, V. S. (2012). Criteria for Pedagogical Investigations Quality. *Pedagogical Investigations*, 1, 51–63.
- Bloemeke, S., Koenig, J. & Busse, A. (2014). No TitleThe Transition from Teacher Training to the Teaching Profession. *Zeitschrift Fur Erziehungswissenschaft*, 17(3), 509–42.
- Burlai, V. V., Suflyayeva, N. E. & Yurenkova, L. R. (2012). Monitoring and Estimation of Students' Knowledge Using Module-Rating System. *Electronic Scientific and Technical Periodical Science and Education*. Retrieved from <http://technomag.edu.ru/doc/323937.html>.



- Elliot, D., Wilson, D. & Boyle, S. (2014). Science Learning via Multimedia Portal Resources. *British Journal of Educational Technology*, 45(4), 571–580.
- Fidalgo-Blanco, Á., Sein-Echaluce, M. L. & García-Peñalvo, F.J. (2014). Knowledge Spirals in Higher Education Teaching Innovation. *International Journal of Knowledge Management*, 10(4), 16–37.
- Golovacheva, V. N. & Abaeva, N. F. (2015). Theory and Practice of Pedagogic Texts Formation. Saarbruecken: LAP LAMBERT Academic Publishing.
- Golovacheva, V. N., Abaeva, N. F. & Kokkoz, M. M. (2015). To the Issue of University Tests for Students. *Higher School of Kazakhstan*, 3, 35–39.
- Ibragimov, I. D., Dusenko, S. V., Khairullina, E. R., Tikhonova, N. V. and Yevgrafova, O. G. (2016). Recommendations on the Textbooks Creation as Information and Teaching Tools of Education Management. *International Electronic Journal of Mathematics Education*, 11(3), 433-446.
- Ji, J. (2013). Human Activity and Social Groups as the Evaluation Natural Habitat: Reflection on the Teaching and Evaluation in XXI Century. *Education Issues*, 1, 73–106.
- Kadnevskiy, V. M. (2012). Francis Galton, One of the Scientific Test Method Creators. *Pedagogical Diagnostics*, 6, 27–41.
- Kamalova, L. A. & Raykova, E. (2016). The Quality and Criteria of Evaluation of Educational Work at the Universities of Russia at the Contemporary Stage. *International Electronic Journal of Mathematics Education*, 11(1), 71-79.
- Kim, D. & Jang, S.E. (2014). Dialogic Practices in Using Podcasting and Blogging as Teaching Tools for Teachers Seeking ESOL Certificate. *Journal of Educational Computing Research*, 51, 205–32.
- Liu, H. J., Lan, Y. J. & Ho, Y. Y. (2014). Exploring the Relationship between Self the Relationship Vocabulary Learning and Web-Based Collaboration. *Educational Technology and Society*, 17(4), 404–419.
- Martinez-Garza, M. M. & Clark, D. B. (2013). Teachers and Teaching in Game-Based Learning. *Information Science Reference*, 1, 53-56.
- Morris, C. & Chikwa, G. (2014). How Effective Are They and How Do Students Engage with Them? *Active Learning in Higher Education*, 15, 25–37.
- Talbi, O., Warin, B. & Kolski, C. (2013). Towards a Support System for Course Design. In *Proceedings of the 5th International Conference on Computer Supported Education*, 449–454. Aachen: SciTePress.
- Tsai, C. Y., Lin, C. N., Shih, W. L. & Wu, P. L. (2015). The Effect of Online Argumentation upon Students' Pseudoscientific Beliefs. *Computers & Education*, 80, 187–97.
- Tsaritsentseva, O. P. (2013). Models of Pedagogical Testing. *Education Technology*, 1, 103–14.
- Van den Hurk, H. T. G., Houtveen, A. A. M., Van de Grift, W. J. C. M. & Cras, D. W. P. (2014). Data-Feedback in Teacher Training. Using Observational Data to Improve Student Teachers' Reading Instruction. *Studies in Educational Evaluation*, 42, 71–78.
- Waight, N., Chiu, M. & Whitford, M. (2014). Factors That Influence Science Teachers' Selection and Usage of Technologies in High School Science Classrooms. *Journal of Science Education and Technology*, 23, 668–81.
- Yarullin, I. F., Prichinin, A. E. & Sharipova, D. Y. (2016). Risk Management of an Education Project. *International Electronic Journal of Mathematics Education*, 11(1), 45-56.
- Zaytseva, T. V., Pusnaya, O. P., Nesterova, E. V., Smorodina, N. N., Igrunova, S. V. & Putivtseva, N. P. (2013). On Developing the Adaptive Knowledge Check Model. *Scientific Journal of Belgorod State University*, 15-1(158), 223–27.
- Zaytseva, T. V., Smorodina, N. N. & Vasina, N. V. (2013). Application of the Expert System of Knowledge Assessment «REXPert» in the Educational Process. *Scientific Journal of Belgorod State University*, 22-1(165), 231–235.