Review of European Studies; Vol. 7, No. 7; 2015
ISSN 1918-7173 E-ISSN 1918-7181
Published by Canadian Center of Science and Education

Specifics of Development of the Integral Method of Knowledge Estimation

Viktoriya Nikolaevna Golovachyova[1], Nella Fuatovna Abayeva[2], Mahabbat Meyramovna Kokkoz[3], Lezzetzhan Muhamedzhanovna Mustafina[2] & Bakhytzhan Muhamedzhanovna Mustafina[4]

[1] The Chair of measuring and computing systems, Karaganda State Technical University, Karaganda, Republic of Kazakhstan

[2] The Chair of Mathematics, Karaganda State Technical University, Karaganda, Republic of Kazakhstan

[3] The Chair of Information Technology and Security, Karaganda State Technical University, Karaganda, Republic of Kazakhstan

[4] The Chair of computer and software engineering, Kazakh National technical university after K. I. Satpayev, the Republic of Kazakhstan

Correspondence: Abayeva Nella Fuatovna, the Chair of Mathematics, Karaganda State Technical University, Karaganda, Republic of Kazakhstan. Tel: 8-721-230-2690.

Received: March 19, 2015 Accepted: April 21, 2015 Online Published: May 14, 2015

doi:10.5539/res.v7n7p284 URL: http://dx.doi.org/10.5539/res.v7n7p284

Abstract

The problem of testing-based objective estimation of knowledge acquires new forms and content in the context of new paradigms. Analysis of current test methods suggests that sometimes questions with answers assuming multiple choice or multiple choice and formulation do not allow objective estimation of students' knowledge that results in reduction of the simulating effect of pedagogical grades on the cognitive activity of students and educational process quality in general. This article suggests an integral method of knowledge estimation based on a new approach to question and answer formulation enabling free formulation of a test answer. The theoretically justified and experimentally verified data can be used in order to improve control and estimation of knowledge by the social and humanitarian subjects.

Keywords: testing, forms of test answer formulation, integral method of knowledge estimation, education quality

1. Introduction

In the XXI century, didactics is oriented to strict control of all stages of educational process from purpose and content development to verification of results. Therefore, pedagogical science actively seeks the ways and means of knowledge control and estimation in order to increase education quality. Scientist K. Ingeskamp thinks that "modern scientifically based didactics is bound to be defeated if it is not based on many tools of maximum objective methods of pedagogic diagnostics."

Naturally, this means objective control, i.e. such knowledge assessment methods and, broadly speaking, pedagogic diagnostics, which enable a teacher or a researcher to use such means that provide accurate and complete information on the knowledge level and quality of the educational process in general. At the present stage of development, pedagogic science considers testing to be such a means.

Testing is targeted examination, which is equal for all testees, performed in strictly controlled conditions and enabling objective estimation of the studied characteristics of the pedagogical process.

Objectivity, i.e. independence of the verification and knowledge assessment from the qualified teacher proficient in this field, is an advantage of this form of control. However, from our point of view, sometimes application of this method does not allow objective estimation of the knowledge level. Herewith, the following question arises: how has the testee managed to answer the test: through logical reasoning or randomly? Besides, physical acquisition of training material is always possible.

Examples indicate that rationally compiled tests include the tests requiring answers in one of the following forms:

1) Selection of the correct and complete answer from a series of the proposed ones (correct or incorrect, complete or incomplete, accurate or inaccurate);

2) Selection comprising of two parts (the first part requires any selection, and the second one requires justification of the selection);

3) Alternative selection (yes/no, 0/1, true/false);

4) Arrangement of elements from the proposed list in the correct succession;

5) Matching elements from two lists;

6) Statement completion, indication, or selection of omitted words;

7) One word (or number) answer;

8) Answer in many words limited in the order or inter-word connections.

The available possible forms of presentation of test answers are appropriate. Their applicability in the training process is justified. It is beyond any doubt that didactic tests are the latest method of control by many characteristics and are definite leaders among traditional forms of knowledge checks.

The form is defined as a communication method for arrangement, adjustment, and existence of the content in general composition of test jobs. The main difficulty of the problem is the contradiction between theoretical and practical reasoning against the form. The majority of practitioners of the test process find the form of test jobs familiar and quite understandable. Therefore, they do not see any problem here. Correspondingly, the practitioners do not see any reasons to change the forms, to learn forms development and methods for development of new test jobs. Such position of practitioners results in testing degradation.

Current computer source environment enables development of tests including multiple-choice, numerical, and formulated answers. Multiple-chose answers are the most commonly used. They are easier to prepare (they do not require multiple samples of correct answers, which are difficult to make complete) and, most importantly, to use. In case of multiple-choice answers, students direct the main efforts to perform their tasks and not to choose answers.

Many years' experience of using testing in the educational process shows that this method has many advantages: audience coverage; simple and efficient verification of results; possible integrated computerization of the testing process; reduction in estimation subjectivity.

However, despite all evident advantages of testing, we have to acknowledge that it also has some disadvantages. These include deep distortion in perception of material integrity by students; stereotyped thinking development; absence of innovative approach; verbatim learning of answers; memorizing and copying of test keys; great dependence of test reliability rate on the variability of test scores of the testees in a certain group, etc.

This situation implies strong negative consequences: reduction in motivating action of grades on the cognitive work of students and the quality of the educational process in general.

Test questions point to ponder. The case is that not every subject can be formalized. Formalization is obvious for such subjects as physics, mathematics, mechanics, and others. Herewith, it is though not always possible to formalize knowledge in social and humanitarian subjects. The main reason for disparity between prospective and operational capabilities of computers is the delay in development of methodological problems and new knowledge evaluation methods.

In order to make test-based knowledge control efficient for the specified subjects, it is necessary to find out the knowledge acquisition level at every educational stage. Herewith, the test tools have to cover all the required characteristics of knowledge acquisition, for example, ability to specify an answer with examples, ability to express own thoughts logically, properly, correctly, etc. Only such forms of knowledge estimation, which are as good in matter as oral control, allow using the great advantages of such an efficient method of accurate and objective estimation of knowledge as testing.

Nevertheless, there is such an important problem as complexity of text answer meaning recognition. As you know, standard methods for system analysis and computer modeling that are based on precision processing of numerical data cannot essentially cover the great complexity of the processes of human reasoning and decision-making. Thus, it is hard to escape a conclusion that significant estimates of humanitarian systems including education require refusal of high standards of accuracy and rigidity, which are usually expected from a mathematical analysis of clearly defined mechanical systems and which provide more tolerate approaches similar by nature.

The testing method can be improved trough a new approach to the formulation of the answer to a test question assuming free text formulation of the answer; the answer can be recognized based on subject knowledge base developed by experts; and the analysis of this answer can be estimated based on a package of its formation criteria, provided the algorithms of estimation criteria for the quality of the answers have been developed.

2. Method

2.1 Definition of the Integral Method of Knowledge Estimation

Integral method is a method enabling objective estimation of students' knowledge by a series of the criteria of its formation. The core of the method is estimation of students' knowledge by test questions assuming free formulation of the answers.

2.2 A Series of Criteria for Estimation of Test Answers

We estimate the answer analysis by the following series of criteria:

● Objectivity. It reflects the basic level of the subject knowledge and is determined by comparison of the correspondence of the applied descriptors with thesaurus descriptors or their synonyms. The quality of the criterion "objectivity" is characterized by coefficient δ. The estimation criterion for δ is the ratio between the number of correctly applied descriptors and the total number of descriptors relevant to every test question in the thesaurus. This criterion is evaluated by the following formula:

$$\sigma = \frac{N}{M} \tag{1}$$

where N is the number of descriptors corresponding to the thesaurus for every question of the test:

M is the total number of descriptors corresponding to the thesaurus for every question of the test:

● Literacy. This criterion is determined by the rules for structuring text documents. In this research, we are restricted with the conditions for application of particular factors, e.g. spelling check determined by the comparison of correspondence of every word in the answer text to the words in the spelling dictionary. The quality of the criterion "literacy" is characterized by the coefficient **γ.** The estimation criterion for **γ** is the ration between the correctly spelled words and the total number of words in the answer text. This criterion is evaluated by the following formula:

$$\gamma = \frac{K}{R} \tag{2}$$

where K is the number of correctly written words in the text of answer; and R is the total number of words in the answer text.

● Presence of examples. An example in the answer text clearly explains the main question of the test. This criterion is determined by the comparison of the correlation of examples used in the answer text with the words in the database of examples or their synonyms. The quality of the criterion "presence of examples" is characterized by the coefficient φ. The estimation criterion for φ is a ratio between the number of correct examples in the text of answer and the total number of examples conforming to their database for every test question. This criterion is evaluated by the following formula:

$$\varphi = \frac{D}{F} \tag{3}$$

where D is the number of examples corresponding to the database of examples or their synonyms for every test question;

F is total number of examples from the database of examples or their synonyms for every question of the test.

● Logic connections between sentences. These connections are determined by the rules of the language for connections between sentences. Quality of the criterion "logic connections between sentences" is characterized by the coefficient μ. The estimation criterion for μ is a ratio between the quantity of logic connections in the answer determined by the answer incidence matrix (Figure 1) and the maximum possible number of logic connections between sentences used in the answer text.

	1	2	…	N
1	1	1	1	1
2		1		1
…			1	1
n				1

Figure 1. Answer incidence matrix

This criterion is evaluated by the following formula:

$$\mu = E / L \tag{4}$$

where E is the number of logic connections determined by the answer incidence matrix.

L is the maximum number of logic connections between the used sentences. It is evaluated by the following formula:

$$L = n \cdot (n-1) - \sum_{i=1}^{n-1} i \tag{5}$$

where n is the number of sentences in the text of answer.

Diagonal elements of the matrix determine the number of sentences in the text of answer. Super diagonal elements of the matrix determine connections between the concerned sentences and subsequent sentences of a text answer.

● Complexity. It characterizes the quality of the test answer in general and is determined by the presence of connections between the concerned criteria: objectivity, literacy, presence of examples, logic connections between sentences. The quality of the criterion "complexity" is determined by the coefficient ŋ. This criterion is evaluated by the following formula:

$$\eta = \frac{\delta+\gamma+\varphi+\mu}{z} \tag{6}$$

Where δ is a coefficient obtained by the criterion "objectivity";

γ is a coefficient obtained by the criterion "literacy";

φ is a coefficient obtained by the criterion "presence of examples";

μ is a coefficient obtained by the criterion "logical connections between sentences";

z is the number of the considered criteria.

The complexity analysis is provided by expertise. The general score for every question of the test is assigned by the following formula:

$$Ball = \left[\frac{\overline{\delta}+\overline{\gamma}+\overline{\varphi}+\overline{\mu}+\overline{\eta}}{w \cdot k}\right] \tag{7}$$

Where $\overline{\delta}$ is a grade assigned by the criterion "objectivity";

$\overline{\gamma}$ is a grade assigned by the by criterion "literacy";

$\overline{\varphi}$ is a grade assigned by the criterion "presence of examples";

$\overline{\mu}$ is a grade assigned by the criterion "logic connections between sentences";

$\overline{\eta}$ is a grade assigned by the criterion "complexity";

w is the total number of criteria of the answer analysis;

κ is a complexity coefficient for every test question.

● Estimation of the content of a "verbal and linguistic" answer. Let us consider an example of verbal and linguistic answer to the question "What is an array of information"?

This question may have one of the following answers:

"An array of information is an information structure consisting of one or more records, so that the records describe objects and the array describes the class of objects. Several arrays form a system or series of arrays. A record is a specified set of data, which characterizes some objects or processes. Record examples are sales vouchers, work orders, invoices, questionnaires, statistic records".

2.3 Series of Oral Answer Criteria

As is obvious, an oral answer is a specific text, the content of which discloses the essence of the raised question. Accordingly, estimation of the answer, first of all, requires estimation of its content by the developed complex of criteria:

● Objectivity. It determines the basic level of the subject knowledge and is determined by the comparison of correspondence of the applied terms to the thesaurus terms. This criterion is a constituent part of the basic vocabulary of the subject.

The basic vocabulary of the subject field is the main reference didactic material used for an analysis of every text answer. It is divided into the following components:

● A thesaurus is a kind of a vocabulary with terms placed in a certain order (the principal sense-bearing items) and connections fixed between them. A thesaurus is used for the search of words determining the basic level of the subject knowledge. Any thesaurus consists of an introduction, an alphabetic index (vocabulary), and a classified catalog.

The main method of thesaurus formation is selection of descriptors, which are usually nouns, from any training module, notes, working programs, synonyms, and other linguistic sources, which are specific for the subject.

● An example database is a vocabulary consisting of words placed in a certain order, which vividly explain the main questions of the test. The main method for formation of the vocabulary is selection of examples provided by teachers at lectures, answers of students at practical studies, seminars, etc.

Example database development consists of two interrelated stages:

1) Initially, it is necessary to form a set of various example words.

2) Then, the most frequently used examples are selected. Moreover, all forms of the example words are considered.

Let us separate descriptors from the text of the considered answer by the morphological analysis of sentences. The analysis of a testee's answer results in division of the answer into sets of various parts of speech: a set of nouns, a set of verbs, a set of adjectives, etc. We need to count the number of nouns in the sentence in order to select descriptors from the text of the considered answer. These are the following words: *array, information, structure, record, object, class, system, series, data, and process.* Herewith, various forms of every word and its synonyms must be taken into account. Consequently, the thesaurus must include the selected words-descriptors and consider all the forms of the corresponding words, including synonyms.

Based on the above, the following settlement procedures for determination of the criterion "objectivity" have been selected:

1) Sentence analysis.

2) Selection of key words (descriptors).

3) Determination of the correspondence with thesaurus descriptors.

4) Quantitative estimation.

5) Score determination.

● Literacy is determined by grammar rules of the text document structure. Quality of the criterion "literacy" is characterized by the coefficient γ. The answer literacy is determined by several characteristics:

$\gamma 1$—presence of a subject; $\gamma 2$—presence of a predicate; $\gamma 3$—presence of objects; $\gamma 4$—presence of attributes; $\gamma 5$—presence of adverbial modifiers; $\gamma 6$—absence of grammar mistakes; $\gamma 7$—absence of errors of style; $\gamma 8$—presence of various syntactic constructions.

The selected coefficients characterize the answer as an overall literacy rate, for example: A is a complete sentence; $\gamma 1 \cap \gamma 2 \cap \gamma 3 \cap \gamma 4 \cap \ldots \gamma k$, where k is the number of structural and semantic components, which determine this rate.

If k = max I, where I is the maximum number of structural and semantic components, which are in compliance with the definite condition, the answer shall be considered as a complete one:

B is an incomplete sentence;

$\gamma 1 \cap \gamma 2 \cap \gamma 3 \cap \gamma\, k$, where K < I;

C is a limited sentence—only $\gamma 1 \cap \gamma 2$;

D is an irregular sentence—there is no γ1 or γ2;

E is a faulty sentence with grammar mistakes;

F is an inaccurate sentence with errors of style.

Series of various rules for evaluation of grammar quality of the text answer in general may be an estimation criterion γ.

- As we know, examples in the text of answer vividly explain the main question. The quality of the criterion "presence of examples" is characterized by the coefficient φ and determined by comparison of the correspondence of the applied examples to the words from the example database. The above answer to the question "What is an array of information?" contains five examples, as it has been specified: sales vouchers, work orders, invoices, questionnaires, statistic records.

- The rules of language connections between sentences determine logic connections between sentences. It is commonly known, that a text consisting of two or more sentences, which are interconnected by sense or structure and have equal function on a compositional or stylistic level, is a complex syntactical unity.

In the context of structure and semantic, complex syntactical unities are divided into complex syntactical unities with chain connection and complex syntactical unities with parallel connection between sentences.

μ may have the following estimation criteria:

- All the sentences of the answer shall have logic connections with the principal sentence of the answer;

- All the sentences, except for the principal one, shall be subdominant to the principal sentence;

- All the sentences shall complete content of the principal sentence, discover its essence in more details, explain the principal sentence, classify subjects of the answer, etc.

We propose to construct an incidence matrix of the text answer in order to formalize this complex criterion.

Diagonal elements of the matrix determine the number of sentences in the text of answer. Super diagonal elements of the matrix determine connection between the concerned sentences and subsequent sentences of the answer, i.e. successive logical connections. In the above answer to the question "What is an array of information?" diagonal elements of the matrix consist of four components, because the answer text consists of four sentences. There is a logic connection between the first and the second sentence. It is determined by the keyword "array." There is a logic connection between the first and the third, the first and the fourth, and the third and the fourth sentences of the answer. It is determined by the keyword "record". The criterion "logic connections between sentences" is specified by the number of logic connections in the answer text.

- Complexity. This criterion characterizes the quality of every text answer in general and is determined by the presence of connections between the concerned criteria: objectivity, literacy, presence of examples, logic connections between sentences. An answer is considered to be complex within the meaning of general characteristic if in the answer to the main question a testee manages to determine criteria "objectivity" at a sufficiently high level, forms a grammatically correct text of the answer, demonstrates examples explaining the substance of the main question, and gradually develops his/her idea in the answer text. The complexity is analyzed by counting the number of connections in the answer and determined by expertise.

General score for every question of the test is assigned by the formula (7).

2.4 Scoring

Grades are assigned with consideration of every selected criterion according to a scale of one to ten. The main problem for final score assignment is determination of the limits between two points of the traditional scale of one to five, when knowledge can be estimated higher or lower than the corresponding grade. Fuzzy sets of criterion grades are applied in order to determine grade limits in the research, and a heuristic approach is applied for determination of final score.

Further, it is necessary to determine the limits of fuzzy assessment of the dedicated parts. Herewith, the main purpose and difficulty is to determine the limit values of the dedicated parts. In case of a normal law for the grade allocation function, the grade its parameters can be used for selection of the grade limits. For example, a property of the normal probability law is a standard deviation σ. We take the function describing the confidence probability as a membership function, and a standard deviation of the normal law of distribution as the criterion for identification of fuzzy set limits. Confidence figures of estimand within the confidential interval are calculated for the standard intervals by standard integrals (Table 2).

Table 2. Grade confidence figures

Interval	Probability
-σ, +σ	0.6827
- 2σ, +2σ	0.9545
-3σ, +3σ	0.9973

For example, if the confidence interval is 2σ, the lower limit of the grade is X_1=M-2σ, and the upper limit is X_2=M+2σ, where M is a parameter of the grade distribution function. Thus, it is valid to say that the score is equal to the considered grade with a probability of 0.9545, and to the previous or following grade with a probability of 0.0227.

The confidence interval is selected based on the result accuracy requirements, and the more accurate results are required the less the value of the confidence interval is.

Similarly, it is possible to determine the limits of the fuzzy grade for all criteria applied in determination of the final score.

Let N grades are provided by the criterion “objectivity” for a certain answer.

Mathematical expectation is calculated by the following formula:

$$M = \sum_{i=1}^{n} \delta_i / N \tag{8}$$

where n is the number of grades; i is the sequence number of the grade.

The result is M =8.

Dispersion is evaluated by the following formula:

$$D = \sum_{i=1}^{n} (\delta_i - M)^2 / N \tag{9}$$

where n is the number of grades; i is the sequence number of the grade; M is mathematical expectation.

The result is D =1.4.

The standard deviation is calculated by the following formula:

$$\sigma = \sqrt{D} \tag{10}$$

where D is the dispersion.

The result is σ = 1.18.

When the grade distribution function parameters are determined, for example, by the criterion "Objectivity": mathematical expectation, dispersion, standard deviation, we can determine the limits of the fuzzy grade depending on the desired accuracy.

When fuzzy sets are applied, a preference function is composed for every grade, the confidence probabilities are selected, the borders of fuzzy sets are determined, and then the grade on a scale of one to five is assigned.

Whereas formation of the final grade in integral method is based on a set of criteria, this method is applied for every criterion.

Pedagogically, the structure of knowledge of every student must be considered at grade assignment. Therefore, we have developed the following heuristic algorithm: every form of the test consists of 10 proposed test questions and numbers of the questions are ranked in difficulty growth. 3 (three) first test questions are simple, the following 5 (five) questions are moderately difficult, and the last 2 (two) questions are difficult.

For the cases of heuristic approach to estimation, we have developed an algorithm, which is a task of combinatorial theory and is formed as follows: it is required to determine various values of α, β, and λ that would meet the conditions for the grade obtaining on a scale of one to five:

$$\begin{aligned} a \le \alpha_k + \beta_m + \lambda_n \le d \ \text{(excellent)} \\ b \le \alpha_k + \beta_m + \lambda_n \le a \ \text{(good)} \\ c \le \alpha_k + \beta_m + \lambda_n \le b \ \text{(satisfactory)} \\ 0 \le \alpha_k + \beta_m + \lambda_n \le c \ \text{(unsatisfactory)} \end{aligned} \quad (11)$$

with the threshold values of k, m, n.

where α, β, λ are various combinations of difficult, moderately difficult, and simple questions, correspondingly;

k is the total number of difficult questions in the test;

m is the total number of moderately difficult questions in the test;

n is the total number of simple questions in the test;

d is the total number of questions in the test;

a, b, c is the number of correct answers meeting the conditions of being assessed on a scale of one to five.

2.5 Specifics of Development of the Expert Control System and Integral Method of Knowledge Estimation (IMKE)

This integral method of knowledge estimation is a base for development of the system of expert control and estimation of knowledge—IMKE.

The core of the system is a base vocabulary of the subject field developed and accumulated in the process of formulation and education of students and which is the main reference didactic material. It is used for analysis of every text answer and consists of the following elements:

● Thesaurus The main method of the thesaurus development is selection of descriptors, which are usually nouns, from lecture notes, a working program, synonyms, and other linguistic sources specific for the subject.

● Example database. The main method for development of the base is selection of examples provided by teachers at lectures, answers of students at the practical studies, seminars, etc.

● Spelling dictionary. Input information of the testee shall be fixed and entered into the lexical analyzer of the expert system IMKE. The lexical analyzer receives the original text of an answer directly from the input interface elements and transforms it into a lexical items array. The analyzer searches for every lexical item using the basic vocabulary of the subject field. Search suggestion shall be successful in case of exact match of the analyzed word with descriptor of the basic vocabulary of the subject area. In this case, the corresponding information is transferred to the program.

The criterion "Objectivity" is calculated based on this information. Similar information is selected and prepared for settlement of other criteria.

The system of control and estimation of knowledge IMKE assigns final test score providing a corresponding result summary based on the tests results for each testee.

An example of the result summary is shown in Table 1 where coefficients enable a detailed analysis of testees'

answers to each question of the test and justify the final test score.

3. Results

Perennial research in the field of knowledge estimation performed by various authors shows that the grades used for estimation of the students' knowledge in the groups shall be distributed under a normal law. Accordingly, the most efficient knowledge estimation system is a system preventing underestimating or shifting average grade in the group answers, i.e. the main operational hypothesis is a hypothesis of normal point allocation in the process of education control.

Using this hypothesis, we have checked the validity of knowledge estimation results with the system of control and estimation of knowledge IMKE. For benchmarking, we selected the theoretically "perfect" knowledge control and estimation system, which was called theoretical. Accuracy and efficiency of the developed method of knowledge estimation can be determined by comparison and theoretical evaluation of the results of knowledge estimation by testing (the control group) and by the method of integral estimation of knowledge (the experimental group). It is expected that the closer the parameters of the investigated system to the theoretical one are, the more perfect the system is.

In order to ensure validity of the results, we tried to identify:

- The interconnection (r) between the considered knowledge estimation methods (the testing method (T) and the method of integral estimation of knowledge (I));

- The statistical characteristics of the estimation results by the type of control: average point (T_{av}; I_{av}), standard deviation (σT; σI);

- Whether the sampled data conform to the hypothesis of normal distribution of the population by the method χ^2 (K. Pierson's criterion) with the significance value equal to 0.05.

The obtained results of statistical processing of experimental data demonstrate that the score distribution functions applied in the control and experimental groups are close and are regulated by the same law. At the same time, grade frequency distribution function of the experimental groups are closer to the theory than the control groups (the probability is lower than the critical values obtained on the data processing in the control groups) ($P_T(\chi^2 \geq \chi_q^2) = 0.0047$; where k = 1 and $x_q^2 = 7.514485$; $0.0047 < 0.005$; $P_I(\chi^2 \geq \chi_q^2) = 0.0833$ where k = 1 and $x_q^2 = 2.985654$; $0.0833 > 0.005$).

Thus, the developed method of integral estimation of knowledge enabled education improvement due to providing teachers with objective information on the level of training material assimilation by students; due to a detailed analysis of the content of knowledge estimation, which increases the interest and motivates to study as is evidenced by the data available to the teacher after testing; and due to purposeful correction of the education process with due consideration of testing results, complexity index selection, and alteration of fuzzy grade boundary on assignment of the final grade.

4. Discussion

The problem of the testing-based objective estimation of knowledge acquires new forms and content in the context of new paradigms. Wide spreading of the method of knowledge estimation resulted in the problem of estimation of students' educational results by the teacher as an independent line of pedagogical science.

An analysis of various approaches to objective estimation of knowledge by testing suggests that, notwithstanding of all results in this field, the issue of the method objectivity, of the sufficiency of the reflection of real knowledge level through the current estimation systems remains open. The problem of knowledge objectivity is determined by the multidimensionality of this issue from the pedagogical, psychological, and methodological points of view.

An analysis of the current condition of the automated system of knowledge control and estimation revealed that the available methods for formulation of questions and answers based on didactic tests (multiple choice questions and questions of the multiple choice and formulation type) have a lot of advantages: they reduce estimation subjectivity; make complex computerization of the knowledge control and estimation process possible; raise productivity; enable simplicity and efficiency of the result verification, etc.

However, examples seem to indicate that available methods for formulation of questions and answers based on didactic tests sometimes prevent objective estimation of the actual knowledge level of students, especially when it comes to the subjects of social and humanitarian cycle. These methods have the following disadvantages: deep distortion in perception of material integrity by students; stereotype development; absence of innovative approach; verbatim learning of answers; memorizing and copying of the test keys; sometimes it is difficult to

find out the way of answer chosen by the student (logical reasoning or randomly), etc.

This situation has negative consequences: reduction in the motivating effect of the grades on the cognitive work of students and the quality of the overall educational process.

The routine control practice using pedagogical tests enables receiving objective information on the specific knowledge level of an individual person and his/her skills, and relate these data with the training tasks in order to enable timely correction of the new knowledge acquisition process. The need for well-developed methods of estimation of the students' knowledge level is constantly implemented in the process of study.

Search for pedagogically efficient ways and development of a method that would allow improving the knowledge control and estimation process to raise the education level remains the main problem of our research purposed for theoretical justification and practical development of the method of integral estimation of the knowledge.

In order to improve the knowledge control and estimation process for the education level increase, we propose to improve the testing method by a new approach to the formulation of answer to the questions of the test assuming free text form of the answer. An answer is recognized based on the subject knowledge base developed by experts. An analysis of such answer is provided by estimating using the series of formation criteria, including algorithms of the quality control estimation. There are definite difficulties. The case is that not every subject can be formalized. Formalization is applicable and obvious for such subjects as physics, mathematics, mechanics, and others. Herewith, it is though not always possible to formalize knowledge in social and humanitarian subjects. In order to make test-based knowledge control efficient for the specified subjects, it is necessary to find out the knowledge acquisition level at every educational stage. Herewith, it is necessary to cover all the required characteristics of knowledge acquisition with test tools. For example, such factors, as ability to make an answer concrete by means of examples, knowledge of the facts, ability to express own thoughts logically, properly, correctly, etc. Only such forms of knowledge estimation, which are as good in matter as oral control, allow using the great advantages of such an efficient method of accurate and objective estimation of knowledge.

Nevertheless, there is such an important problem as complexity of the text answer meaning recognition. Standard methods for system analysis and computer modeling that are based on precision processing of numerical data cannot essentially cover the great complexity of the processes of human reasoning and decision-making. Thus, it is hard to escape a conclusion that significant estimates of humanitarian systems including education require refusal of the high standards of accuracy and rigidity, which are usually expected from a mathematical analysis of clearly defined mechanical systems and provide more tolerate approaches that are similar by nature. Possibly, only these approaches will make computer modeling a really efficient method of humanitarian system analysis.

Naturally, innovations are associated with a certain risk due to difficulty in specification of the final result and avoidance of erroneous assumptions. Achievement of the qualitative result from innovation requires mature reflection, detailed analysis, and competent arrangement.

Accuracy and objectivity of the knowledge estimation depends not only on formulation of text answers or on the underlying criteria and what the designated factors of student knowledge estimation are in particular, but also on what the grade scale or system is.

5. Conclusion

This article suggests a method of integral estimation of knowledge based on a new approach to the question and answer formulation enabling a free formulated form of test answers. Theoretically justified and experimentally verified data can be used in order to improve control and estimation of knowledge in the subjects of social and humanitarian cycle.

In the present context, the analysis of theory and practice of the knowledge control and estimation let us make the following conclusions:

1) Currently, the automated systems of knowledge control and estimation based on didactic tests and various approaches to assignment of grades are widely used for verification of the process of knowledge acquisition. Dedicated constructions of the test questions and answers: multiple choice or multiple choice and formulated questions do not always provide objective estimation of the student knowledge, especially on the social and humanitarian subjects. This situation has strong negative consequences: reduction in motivating effect of the grades on the cognitive work of students and the quality of the educational process in general.

2) Currently, there are no automated systems of knowledge control and estimation, which would enable assessment of such factors as ability to specify an answer with examples, demonstrate awareness of facts, show

the skill of expressing oneself logically and correctly, etc. The issue of the need for a system of knowledge control and estimation, which would enable estimation of the real level of students' knowledge on the social and humanitarian subjects, where knowledge and reasoning play the main role, is especially acute.

3) The main reason for disparity between the prospective and operational capabilities of computers is a delay in the development of methodological problems and new methods of knowledge estimation.

4) In the process of consideration of the problem of improvement of the knowledge control and estimation to improve the education quality, we have examined the ways and means to improve this method. There is a theoretically justified need to apply a new approach to formulation of test answers assuming a free formulated form of the test question and answer and also a need to develop criteria for analysis of this answer and scientifically based approach to estimation.

Thus, using the developed integral method of knowledge estimation, we managed to achieve the education quality improvement through:

- Providing teachers with objective information on the level of the training material assimilation by students—the shape and nature of the curve of estimation density function match;

- A detailed analysis of the estimation content, which increases the interest in and motivation for studying as is evidenced by the data available to a teacher after testing (table 1);

- Purposeful correction of the education process with due consideration of testing results, complexity index selection, alteration of fuzzy grade limits at forming the final score.

- An analysis of the theory and practice of current knowledge control and estimation status has allowed justifying theoretically the need to develop a method of integral estimation of knowledge;

- A search of the ways and means of knowledge control and estimation enables development of a model of integral estimation of the knowledge;

- An integral method of knowledge estimation has been developed, which is based on a new approach to the question and answer formulation based on didactic tests that allow freely formulated form of test questions and answers;

- An automated system of the knowledge control and estimation imke has been developed based on the integral method of knowledge estimation;

- The research results, which are theoretically justified and experimentally verified, can be used in order to improve control and estimation of knowledge to improve the quality of education, which proves the suggested hypothesis.

The following methodical recommendations can be formulated on the basis of the conclusion of the practical and experimental work: the automated system of knowledge control and estimation IMKE, which is based on the developed integral method of knowledge estimation, can be applied in the process of teaching students and senior high-school pupils, especially with regard to social and humanitarian subjects;

The obtained results of the research cannot settle all issues of the problem of the quality of knowledge acquired in the process of education. Further theoretical and practical development of this method requires addressing such issues as improvement of the integral method of knowledge estimation in terms of expansion in the number of knowledge estimation criteria, development of criteria of their quality assessment; development of the knowledge base, involving other analyzers, etc.

References

Berk, R. A. A. (1984). *Cuide to Criterion—Referenced Test Construction* (pp. 231-266). Baltimore: The John Hopkins Univ. Press.

Binet, A., & Simon, T. H. (1936). *The Development of Intellegence in Young Children* (pp. 45-79). Vineland, NY: The Training School.

Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F. M. Lord, & Novick (Eds.), *Statistical Theories of Test scores* (Ch. 17-20, p. 568). Reading Mass.: Addison-Wesley.

Bloom, B. S. et al. (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals. In *Handbook 1: Cognitive Domain* (P. 187-215). NY: David Mc Kay Co.

Brownlee, K. A. (1960). *Statistical Theory and Methodology in Science and Engineering*. NY: John Wiley.

Curriculum Frameworks for Mathematics and Science. (1993). In D. Robitaille (Ed.), *TIMMS Monograph* (No. 1,

p. 616). Vancouver Canada: Pacific Educational Press.

Department of Education and Science. (1989). *National Assessment: The APU Science Approach*.

Gagne, R. M. (1977). *The Conditions of Learning* (3rd ed., P. 149-162). NY: Holt, Rinehart and Winston.

Gardner, E. S., Jr. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, *4*, 1-28. http://dx.doi.org/10.1002/for.3980040103

Golovachyova, V. N. (2001). *Osnovi primeneniya testjvih tehnologiy v pedegogike*. Almati: Izdatelstvo "Gilim".

Gorbunov, A. V. (1998). *Algoritmi testirovaniya i metodi kontrolya Almati: "Avtomatika i informatika"* (No. 1, C. 72-74).

Ingenkamp, K. (1991). *Pedagogicheskaya diagnostika*. M.: Pedagogika.

Keeves, J. P. (Ed.). (1988). *Educational Research Methodology, and Measurement: An International Handbook* (p. 620). NY: Perg. Press.

Kelley, T. L. (1927). *Interpretation of Educational Measurements* (p. 363). NY: World Books Co.

Kingsbury, G. G., & Zara, A. R. (1991). A comparison of procedures for content—Sensitive item selection in computerized adaptive tests. *Applied Measurement in Education*, *4*, 241-261. http://dx.doi.org/10.1207/s15324818ame0403_4

Linn, K. L. (1989). *Educational Measurement* (3rd ed.). NY: Macmillan.

Lord, F. M. (1980). Application of Item Response Theory to Practical Testing Problems. *Hillsdale* (p. 266). NY: Lawtrence Erlbaum Ass. Publ.

Lord, F. M. (1970). Estimating Item Characteristic Curves Without Knowledge of Their Mathematical Form. *Psychometrics*, *35*, 43-50. http://dx.doi.org/10.1007/BF02290592

Lord, F. M., & Novick, M. (1968). *Statistical Theories of Mental Test Scores* (p. 568). MA: Addison Wesley Publ & Co. Reading.

Measuring student knowledge and skills. (1999). *A New Framework for Assessment*. OECD.

Nievesm, S. B. (1999, June 9-11). Quality and Innovation in Education. In *European Quality Congress* (43rd ed., pp. 217-219). Proceeding Book, Forum CALIDAD, Madrid.

Rasch, G. (1997). On Specific Objectivity: An Attempt of Formalizing the Generality and Validity of Scientific Statements. *Danish Yearbook of Philosophy*, *14*, 58-94.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests, Copenhagen, Denmark* (p. 216). Danish Institute for Educational Research, Munksgaard, Copenhagen.

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

The PISA 2003 Assessment Framework—Mathematics, Reading, Science and Problem Solving Knowledge and Skills, OECD. (2003).

Thorndike, R. L. (1982). *Applied Psychometrics*. Boston: Houghton. Miffli Co.

Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measuremen*t, *21*, 361-375.

Yager, R. R. (1986). *Necheytkie mnozhestva i teoriya vozmozhnostey*. Poslednie dostizheniya. M.: Radio isvyaz. http://dx.doi.org/10.1111/j.1745-3984.1984.tb01040.x