

Research Article

Multicriteria Model of Students' Knowledge Diagnostics Based on the Doubt Measuring Level Method for E&M Learning

V. V. Serbin ¹, K. S. Duisebekova,^{1,2} A. M. Smayl ¹, L. M. Alimzhanova,¹ G. D. Zakirova,¹
A. N. Moldagulova ¹, R. J. Satybaldiyeva ¹ and Y. S. Maulenov ¹

¹International Information Technology University, Almaty, Kazakhstan

²Al-Farabi Kazakh National University, Almaty, Kazakhstan

Correspondence should be addressed to Y. S. Maulenov; elzhanms@gmail.com

Received 20 December 2018; Revised 4 April 2019; Accepted 23 April 2019; Published 2 June 2019

Academic Editor: Filippo Sciarrone

Copyright © 2019 V. V. Serbin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This research is devoted to the development of a multicriteria model, valid only for testing, which takes into account the number of responses, the level of question complexity depending on the time spent on their response, time of the whole testing, doubts, number of passes, and the use of additional programs. Also the developed method of measuring the level of the user doubts is described, which gives a clearer and “transparent situation picture” for a more objective decision-making. The method makes it possible to reduce the probability of guessing the correct answer, which increases the objectivity of assessing the knowledge level in the diagnostic systems for E&M learning.

1. Introduction

Distance education and e-learning have laid the foundation for a new global phenomenon, Smart Education, which is not so much about technology as it is about the new philosophy of education.

Developers of distance learning systems cannot always fully take into account all the laws of the learning process and therefore use not quite correct models. And even the most successful solutions in the field of online training rarely get widespread.

Undoubtedly, distance learning is by no means a complete substitute for traditional training, because it is difficult to replace live communication with a teacher or the educational atmosphere that develops in the classroom between subjects of the learning process. At the same time, modern ICT (information and communication technologies) can minimize the “narrow” places of distance learning. Modern pedagogy is also changing, and nowadays transfer of part of the training load to the online mode is quite an acknowledged educational scenario. This can be done when developing educational content on a modern IT basis, for example, using MOOCs (a mass open educational course).

Today, we can confidently state the existence of a new digital Z-generation of people for whom a mobile phone, a computer, and the Internet are natural elements of their living space. In modern conditions, effective education is education without reference to time and place. This education teaches through everyday life. The transition to this mobile learning technology involves the application of new methods, approaches, and principles of the learning process organization. When developing content for mobile devices, it is necessary to take into account that it is intended for young people of the Z-generation, as the method of preparing training content for them differs from the traditional learning content.

So, smart technologies in education, such as mobile applications, are of great importance as they allow to optimize the costs of the university logistics and also to raise the quality of educational services and products to a new level. It is smart technology that allows to develop revolutionary teaching materials, as well as to form individual trajectories of training.

In 2015, UNESCO published Recommendations on Mobile Learning Policies [1], which fully justifies the need for the introduction of mobile technologies in the

educational process. According to these recommendations, “In a world in which dependence on means of communication and access to information is growing, mobile devices will not be a transient phenomenon. As the capacity and capabilities of mobile devices are constantly growing, they can be used more widely as educational tools and take a central place, both in formal and informal education.”

History shows that the competitiveness of the national economy as a whole depends on the development of information technologies. According to the World Economic Forum, the competitiveness index of the economies of states has a high level of correlation with the countries’ information and communication technologies. According to the 2012 rating of the competitiveness of 142 countries of the world by the World Economic Forum, countries that are actively developing information technologies are ahead of Kazakhstan which ranks 51st in terms of creating demand for information technology (the USA-13, Germany-19, India-63, and Egypt-96) and occupies the 104th place in the information technology business conditions place (USA-21, Germany-38, and India-72).

Scientific novelty consists in developing a method of objective measurement of students’ knowledge on the basis of a multicriteria decision-making model.

2. Development of a Multicriteria Model for Diagnosing Knowledge

To solve the problem, an intellectual multicriteria model for assessing students’ knowledge has been developed, which makes it possible to take into account the following parameters: (1) time spent on the answer, (2) the number of correct answers, (3) time of all testing, (4) the doubts of the test person, (5) omission of questions, (6) use of additional programs, and (7) psychological characteristics of the tested person.

Given in the following sections are the coefficients of the model and the methods for measuring the coefficients.

2.1. Coefficient of Accounting for Time Spent on the Answer. Tests contain questions of different levels of complexity (50% of them are easy (level a), 30% are average (level b), and 20% are difficult (level c)). Therefore, the score for each correct answer should depend on the complexity of the question, as well as on the time spent on it.

The maximum score G_{\max} (subject to a quick response in less than 15 s) can be calculated from the formula

$$G_{\max} = A \cdot a_{\max} + B \cdot b_{\max} + C \cdot c_{\max}, \quad (1)$$

where A, B, C is the number of questions on levels of complexity a, b, c , a_{\max} is the maximum score for answering easy questions (level a), b_{\max} is the maximum score for answering average questions (level b), and c_{\max} is the maximum score for answering difficult questions (level c).

$$A + B + C = V, \quad (2)$$

where V is the number of questions in the test.

The maximum score G_{\max} can be typed with a different number of questions in the test.

Points scored for the elapsed time G_{cost} can be calculated by the formula

$$G_{\text{cost}} = \sum_{i=1}^A a_i + \sum_{i=1}^B b_i + \sum_{i=1}^C c_i, \quad (3)$$

where a_i refers to the points for the correct i -question level a , b_i refers to the points for the correct i -question level b , and c_i refers to the points for the correct i -question level c .

The coefficient K_1 characterizing the share of knowledge at score “5” under the condition of a quick response (for a time not exceeding 15 seconds) is determined by the formula

$$K_1 = \frac{G_{\text{cost}}}{G_{\max}}, \quad (4)$$

where G_{cost} is the number of points scored for the elapsed time, subject to the choice of the correct answer.

Provided that all the answers are correct and the time for making a decision of each answer will not exceed 15 seconds, the knowledge coefficient of the tested “excellent” is 1.

2.2. Coefficient of the Number of Correct Answers

$$K_2 = \frac{N_{\text{answers}}}{V}, \quad (5)$$

where K_2 is the coefficient that characterizes knowledge of all issues regardless of time, N_{answers} is the number of correct answers, and V is the total number of questions.

2.3. Coefficient for Time of All Testing. To take account of the time spent on the test (regardless of anything), we need to use the following formula:

$$K_3 = 1 - \frac{T_{\text{cost}}}{T_{\max}}, \quad (6)$$

where K_3 is the coefficient characterizing the passage of the test, T_{cost} is the time spent on testing, and T_{\max} is the maximum allowable time that can be spent on testing.

2.4. Coefficient of Doubt. The mathematical description of doubt as a significant component of the rational choice model, based on the axioms of independence, transitivity, convexity, and monotony of individual preferences, is represented in some way in the study, as discussed by Vinogradov and Kuznetsov [2]. A person builds a sequence of conclusions about the adequacy of his ideas in accordance with his subjective level of conviction. He is convinced of the adequacy of his views in the situation of choosing the type of Ω with respect to goal G , if he believes that the choice of the mode of action C on their basis will allow it to be achieved. At the same time, he perceives some of the characteristics of the X situation Ω ; with respect to the other part, he makes assumptions and shows an intention to prove (to verify their plausibility).

The assumption is the default value of the observed characteristic or a description of the cause-effect relationship between the observed characteristics.

Representations of a person are characterized by a level of conviction. The level of his belief in his ideas about the situation of choosing type Ω with respect to goal G is determined by the frequency of its achievement when choosing the mode of action C based on them. The assessment of the level of conviction changes from zero to one. If the number of unsuccessful attempts to reach goal G when choosing the mode of action C based on human representations increases, then the level of conviction of a person decreases (and vice versa), which becomes an incentive for him to apply efforts for their modification or complete reconstruction due to the growing doubt in the plausibility of the assumptions made. The desire to verify the correctness of assumptions is a measure of a person's doubts.

Thus, the efforts that a person expends to prove (refute) assumptions characterize the degree of his doubts about the ideas about the situation of choosing type Ω as he strives for goal G . According to the provisions of the theory of behavioral psychology, if the level of conviction that depends on the number of confirmations of the correctness of the choice based on representations increases, then the person's desire for verification falls, since he does not see the point in this. The increase in the degree of doubt is an incentive for finding additional arguments (counterarguments).

A parameter that takes into account these two characteristics is the degree of conviction considered as

$$S_U = U_u^\alpha \cdot (1 - S_S)^\beta \in [0, 1], \quad \alpha + \beta = 1, \quad (7)$$

where S_U is the degree of conviction; U_u is the level of conviction (past experience); S_S is the degree of a person's doubt about the correctness of his ideas about the situation of choice; α and β are the coefficients of significance that a person gives to his experience and the need to find evidence. Thus, from the preceding expression, one can derive a formula expressing the degree of doubt.

The degree of doubt in modern tests, in our opinion, is a latent parameter and can be measured only indirectly. If the degree of doubt S_S and the degree of confidence U is estimated on a 100-point scale and expressed in %, then the formula for the degree of doubt S_S will acquire the following form:

$$S_S = 100\% - U. \quad (8)$$

Quantitative characteristics are needed to measure latent parameters. The quantitative parameters for measuring the degree of doubt of the user are the following variables: the number of missed operations, the amount of unconfirmed information, the amount of heterogeneous information (ambiguous from the first time), and the state of the logical chain (the sequence of actions, levels of complexity, etc.).

To determine the level of doubt, we used a superposition of n models, each of which determines the latent parameter of doubt on the quantitative parameters.

Thus, we can postulate the following:

- (1) At the initial level of complexity, the measured level should be the lowest level of doubt.
- (2) On the average level of complexity, the level of doubt should be greater for the advanced level, but less for the initial level.

The quantitative parameters for measuring the level of doubt of the user are the following:

- (i) The number of missed operations
- (ii) The amount of unconfirmed information
- (iii) The amount of heterogeneous information (ambiguous from the first time)
- (iv) The state of the logical chain (the sequence of actions, levels of complexity, etc.)

To take into account the user's level of doubt, the following method is suggested:

- (1) A test is organized, containing i -questions, under the condition $i > 0$.
- (2) Each i -question contains j variants of answers under condition $2 < j \leq 5$.
- (3) The testing questions are divided into training elements. Each training element contains questions of S different levels of complexity. The question contains only one correct variant G_i for the i -question, $G_i > 0$, $G_i \geq n_{ij}$. Each j variant can be preselected and subsequently confirmed with the only accepted variant of the answer k_{ij} under the condition $j = 1$. Each variant of the question has the ability to switch m times. Each question contains a p_{ih} ability to skip an i -question only once, that is, $P = i/p_{ih}$. Then, the amount of heterogeneous information (ambiguous from the first time) can be determined by the number of switching options $m \leq j(j-1)$.

Consequently,

$$D_i = \frac{\phi_1 S_{rw} + \phi_2 S_{wr} + \phi_3 S_{ww}}{m}, \quad \text{under } \phi_1 > \phi_2 > \phi_3, \quad (9)$$

where D_i is the coefficient per doubt of i -question, m is the number of variants of answers in the question, ϕ_1 is the coefficient that characterizes the weight of switching from the correct to the wrong answer, ϕ_2 is the coefficient that characterizes the weight of switching from wrong to correct answer option, ϕ_3 is the coefficient that characterizes the weight of switching from wrong to wrong answer option, S_{rw} is the number of switches from right to wrong answer, S_{wr} is the number of switching from wrong to the correct answer, and S_{ww} is the number of switching from wrong to wrong answer option.

Let us find the average coefficients of doubt $S_{a_{avg}}, S_{b_{avg}}, \dots, S_{s_{avg}}$ to questions of level a, b, \dots, s , respectively,

$$S_{a_{avg}} = \frac{\sum_{i=1}^{n_a} D_i}{n_a}, S_{b_{avg}} = \frac{\sum_{i=1}^{n_b} D_i}{n_b}, \dots, S_{s_{avg}} = \frac{\sum_{i=1}^{n_s} D_i}{n_s}, \quad (10)$$

where s is the number of difficulty levels, n_a is the number of level a questions, n_b is the number of level b questions, and n_s is the number of level s questions.

The final coefficient, which characterizes the average level of doubt of the learning element, is calculated by the formula

$$C_n = \frac{S_{a_{\text{avg}}} + S_{b_{\text{avg}}} + \dots + S_{s_{\text{avg}}}}{s}, \quad (11)$$

where s is the number of difficulty levels.

When $S = 3$ (3 levels: initial, intermediate, and advanced) of each educational element and states 2 (1, answered; 0, did not answer), there are $2^3 = 8$ various states of the logical chain, each of which has its own level of doubts in the matrix states of the level of complexity logical chain. The level of doubt in the matrix depends on the subject area, compiled by the expert.

The weight of the doubt level at $s=3$ for this model corresponds to A -50%, B -30%, and C -20%. The user is given a portion of information of A , B , and C levels of this training element. After that at Table 1, the truth table is analyzed, and the decision is made.

Amount of unconfirmed information is F . If a negative answer was received to the question, then this question is randomly formed even m times during the check-up phase. If the answer is correct the second time, then the chain is simulated a second time, taking into account the level of complexity.

This model defines a latent parameter that is a response to a decision taking into account the level of complexity. The essence of the model is to confirm the doubt of the decision taken by comparing it with a decision of this type.

The latent parameter of this model determines the following quantitative characteristics:

- (i) Subjects
- (ii) Complexity
- (iii) Decision

We propose the following method. Testing of x questions is organized. All x questions are divided into t -groups on a certain topic. Each question has one correct answer.

In testing, all x questions are divided into n complexity levels. Each question has its own level of complexity. The proportion of questions on levels of complexity must satisfy the following condition: for $n = 3$ easy-50%, medium-30%, and complex-20%. Each level of complexity corresponds to the weighting factor F (F_{max} -the most complex, $F_{\text{max}-1}$ -less complex, etc.).

Doubt questions are exposed to which the correct answer was given at all levels of complexity, except the first. Doubts when answering a question of level F are calculated from the answers to questions of the same subject (i.e., the same group) whose complexity is lower than the level of complexity of the current question. Figure 1 shows graph of the dependence of the coefficient K_i of the weight of doubts on the current level of complexity of the question (logarithmic).

TABLE 1: Matrix of states of the logical chain of the complexity level.

State		Doubt level of logical chain L (%)	
A	B	C	
0	0	0	0
0	0	1	80
0	1	0	20
0	1	1	50
1	0	0	20
1	0	1	30
1	1	0	20
1	1	1	0

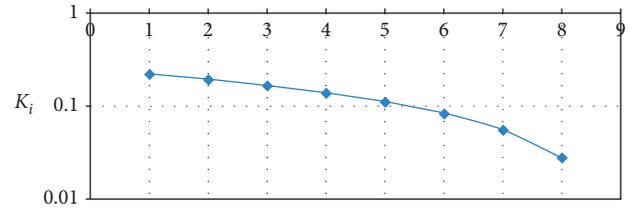


FIGURE 1: Graph of the dependence of the coefficient K_i of the weight of doubts on the current level of complexity of the question (logarithmic).

$$S_F = \sum_{i=1}^m k_i W_i, \quad (12)$$

where m is the number of difficulty levels below that for which the calculation is carried out:

$$m = F - 1, \quad (13)$$

where F is the weighting factor of the current difficulty level; k_i is the weight of doubts per question below the current level:

$$k_i = -\frac{i - m - 1}{\sum_{j=1}^m j}. \quad (14)$$

W_i is the total weight of doubt level i -th question:

$$W_i = \frac{1}{g} \sum_{l=1}^g V_i, \quad (15)$$

where g is the number of questions of the i -th level of complexity of the same subject as the question for which the doubt is calculated; V_i is the decision taken (the answer to the question).

$$V_i = \begin{cases} 0, & \text{if correct answer given,} \\ 1, & \text{if incorrect answer given.} \end{cases} \quad (16)$$

Proceeding from the fact that there can be several questions of the same level of complexity in the test, there is a need to find the mean-square value of the doubt at each level:

$$S_s = \frac{\sqrt{\sum_{t=1}^q S_{Ft}^2}}{q} \quad (17)$$

The final coefficient characterizing the degree of doubt is calculated by the formula:

$$S = \sum_{F=2}^{F_{\max}} f_{F-1} \cdot S_s, \quad (18)$$

where S_s is a doubt in the answers to the question of complexity F and f_F is the doubt weight factor in the answers to the question of the complexity of F :

$$f_F = \frac{F}{\sum_{j=1}^m j}, \quad (19)$$

where F_{\max} is the maximum weight of the highest level. Figure 2 shows graph of dependence of the coefficient f_F of total doubts on the current level of complexity of the question (logarithmic).

Despite the fact that this method uses a small number of characteristics (subject, complexity, and decision) for the calculation, it gives an estimate of the doubts of correctly accepted decisions, which is unquestionably significant for the interpretation of the final test result.

For understanding the developed method of calculating the level of doubt, it is suggested to consider the following example with specific input data.

Let the number of questions = 10, the number of levels $F=3$, the coefficients of the complexity levels of the question be A (easy) = 1, B (medium) = 2, and C (complex) = 3. Correct answers of the user in difficulty levels: $N_A=4$, $N_B=2$, and $N_C=1$. It is required to determine the coefficient of doubt on the average and complex levels of complexity of the questions on which the correct answers were given.

Solution:

- (1) The number of questions from 3 levels of complexity (according to the method) must satisfy the condition:

Questions of the level of complexity $A=50\%$, level $B=30\%$, and level $C=20\%$

Then,

Number of questions of level A (simple) = 5

Number of level B questions (averages) = 3

Number of questions of level C (complex) = 2

- (2) Calculation of the weight coefficients of doubts of the level C in accordance with formula (13):

$$M = F - 1 = 3 - 1 = 2$$

The weight of doubt for level A in accordance with formula (14):

$$k_{1A} = -\frac{i-m-1}{\sum_{j=1}^m j} = -\frac{1-2-1}{1+2} = \frac{2}{3} \quad (20)$$

The weight of doubt for level B in accordance with formula (14):

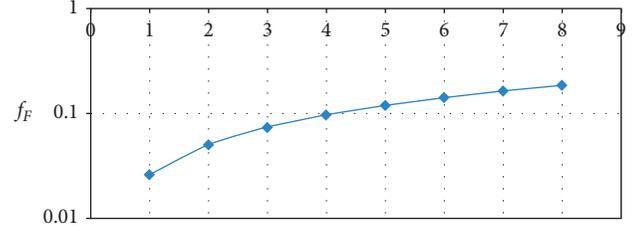


FIGURE 2: Graph of dependence of the coefficient f_F of total doubts on the current level of complexity of the question (logarithmic).

$$k_{2B} = -\frac{i-m-1}{\sum_{j=1}^m j} = -\frac{2-2-1}{1+2} = \frac{1}{3} \quad (21)$$

- (3) Calculation of the total weights of doubt for level C :

The user answered $N_A=4$ out of 5 questions, $N_B=2$ of 3, $N_C=1$ of 2 questions, means number of questions on which the wrong answer was given:

$$g_A = 5 - 4 = 1, \quad g_B = 3 - 2 = 1, \quad g_C = 2 - 1 = 1$$

Let us find the total weight of doubts of level A ($i=1$) in accordance with formula (15):

$$W_i = \frac{1}{g} \sum_{l=1}^g V_l = \frac{1}{5} \quad (22)$$

Let us find the total weight of doubts of level B ($i=2$) in accordance with formula (15):

$$W_i = \frac{1}{g} \sum_{l=1}^g V_l = \frac{1}{3} \quad (23)$$

- (4) Calculation of doubt on the level C questions in accordance with formula (17):

$$S_{3(c)} = \sum_{i=1}^m k_i W_i = \frac{2}{3} \cdot \frac{1}{5} + \frac{1}{3} \cdot \frac{1}{3} = 0.24. \quad (24)$$

Percent of doubts on the issue of the C level = 24%

- (5) Calculation of the weight coefficients of doubts of level B in accordance with formula (13):

$$M = F - 1 = 2 - 1 = 1$$

The weight of doubt for level A in accordance with formula (14):

$$k_{1A} = -\frac{i-m-1}{\sum_{j=1}^m j} = -\frac{1-1-1}{1} = 1. \quad (25)$$

- (6) Calculation of the total weights of doubt for level B :

Let us find the total weight of doubts of level A ($i=1$) in accordance with formula (15):

$$W_i = \frac{1}{g} \sum_{l=1}^g V_l = \frac{1}{5} \quad (26)$$

- (7) Calculation of doubts on the level B questions in accordance with formula (17):

$$S_{3(c)} = \sum_{i=1}^m k_i W_i = 1 \cdot \frac{1}{5} = 0.2. \quad (27)$$

Percentage of doubts about the level of $B = 20\%$

- (8) The meaning of doubt in accordance with formula (19) = 20%

Answer: Doubts arose in 3 correct answers of the student, because questions are raised only on the average (B) and complex (C) levels of complexity.

- (1) In 2 responses of level B , where the correct answers were given, the level of doubt is 20%.
- (2) In the question of level C , where the correct answer was given = 24%.

2.5. Coefficient of Missing Questions (Savings/Waste of Time). If the test program is able to skip the question, then it is logical to keep a record of the gaps that affect the probability of nontripping:

$$K_5 = 1 - \left(\frac{1}{2} \frac{P_a}{P_{a_{\max}}} + \frac{1}{3} \frac{P_b}{P_{b_{\max}}} + \frac{1}{5} \frac{P_c}{P_{c_{\max}}} \right), \quad (28)$$

where K_5 is a coefficient characterizing nontransmissions and P_a , P_b , and P_c are the number of passes in the levels a , b , and c .

In order to keep the coefficient correctly recorded, it is necessary to enter restrictions, the maximum permissible number of passes in the levels a , b , and c .

$P_{a_{\max}}$, $P_{b_{\max}}$, and $P_{c_{\max}}$ are the maximum number of omissions in questions on levels a , b , and c . Provided that $P_{a_{\max}} \leq 15$, $P_{b_{\max}} \leq 9$, $P_{c_{\max}} \leq 6$.

2.6. Coefficient of Using Additional Programs. If, for example, there are problems in testing that require mathematical calculations, then it is necessary to include the "Calculator" in the testing interface. Thus, keep a record of the amount of use of the calculator:

$$K_6 = \frac{\min(L_{\text{used}}, L)}{\max(L_{\text{used}}, L)}, \quad (29)$$

where K_6 is the utilization rate of additional programs, L is the number of tasks requiring calculation, and L_{used} is the amount of use of the calculator.

In this example, the calculator is a useful element of encouragement. There are undesirable but accessible elements of use-hints.

The final formula for calculating the total coefficient characterizing knowledge on "5" is as follows:

$$K = a_1 K_1 + a_2 K_2 + a_3 K_3 + a_4 K_4 + a_5 K_5 + a_6 K_6, \quad (30)$$

where $a_1 K_1$ is the coefficient of knowledge which is excellent under the condition of rapid response, $a_2 K_2$ is the coefficient of knowledge of all questions regardless of time, $a_3 K_3$ is the

test passing ratio, $a_4 K_4$ is the coefficient of confidence, $a_5 K_5$ is the nonrolling ratio, and $a_6 K_6$ is the coefficient of using additional programs.

It was experimentally deduced that $a_1 = 0.23$, $a_2 = 0.5$, $a_3 = 0.07$, $a_4 = 0.1$, $a_5 = 0.07$, $a_6 = 0.03$, and $(a_1 + a_2 + a_3 + a_4 + a_5 + a_6) = 1$.

By calculating the coefficient K , the coefficient of knowledge for "excellent," we can compile Table 2 and determine the assessment.

The final formula for K :

$$\begin{aligned} K = & a_1 \left(\frac{\sum_{i=1}^A a_i + \sum_{i=1}^B b_i + \sum_{i=1}^C c_i}{A \cdot a_{\max} + B \cdot b_{\max} + C \cdot c_{\max}} \right) \\ & + a_2 \left(\frac{N_{\text{answers}}}{A + B + C} \right) + a_3 \left(1 - \frac{T_{\text{cost}}}{T_{\text{max}}} \right) \\ & + a_4 \left[1 - \left[\left(\frac{1}{2} \cdot \frac{S_a}{S_{a_{\max}}} \right) + \left(\frac{1}{3} \cdot \frac{S_b}{S_{b_{\max}}} \right) + \left(\frac{1}{5} \cdot \frac{S_c}{S_{c_{\max}}} \right) \right] \right] \\ & + a_5 \left[1 - \left[\left(\frac{1}{2} \cdot \frac{P_a}{P_{a_{\max}}} \right) + \left(\frac{1}{3} \cdot \frac{P_b}{P_{b_{\max}}} \right) + \left(\frac{1}{5} \cdot \frac{P_c}{P_{c_{\max}}} \right) \right] \right] \\ & + a_6 \left(\frac{\min(L_{\text{used}}, L)}{\max(L_{\text{used}}, L)} \right). \end{aligned} \quad (31)$$

The multicriteria model is valid only for testing, which takes into account the number of answers, the level of complexity of questions depending on the time spent on their response, the time of all testing, doubts, the number of omissions, and the use of additional programs.

3. Experimental Part

3.1. Setting the Experiment. To test the developed model, a large-scale experiment was conducted during two academic years. The experiment involved about 1000 students of 29 courses. All students were divided into 2 groups.

All students took midterm control tests (Att), which the teacher administered and evaluated during the semester.

The first experimental group of students (500 people) passed the final exam (Exm) at the end of the course in the format of computer testing (5 answers and 1 correct one). For the first group, the final grade (F) for the course is calculated using

$$F = 0.6 * \text{Att} + 0.4 * \text{Exm}. \quad (32)$$

The second experimental group of students (500 people) at the end of the same courses passed the exam in the format of computer testing on the basis of the multicriteria model (M), proposed in this study. The results of their average estimates (M) are given in Table 3.

As a result of the experiment, it is required to determine the quality of the developed multicriteria model and its suitability for an objective assessment of the level of knowledge.

TABLE 2: Evaluation depending on K .

Condition	Evaluation	Rating
If $0 \leq K > 0.6$	2 (unsatisfactory)	$K * 100$
If $0.6 \leq K > 0.8$	3 (satisfactorily)	
If $0.8 \leq K > 0.9$	4 (good)	
If $0.9 \leq K \leq 1$	5 (excellent)	

TABLE 3: Experimental data.

No course	Average estimation of intermediate control, Att	Exam estimation, Exm	Final estimation, F	Model final estimation, M
1	68.20	84.30	74.64	71.50
2	66.90	67.80	67.26	69.80
3	51.80	58.40	54.44	55.80
4	53.80	66.30	58.80	59.50
5	78.90	93.00	84.54	80.20
6	81.00	91.00	85.00	75.00
7	55.20	78.40	64.48	65.00
8	75.90	85.80	79.86	80.10
9	87.90	88.10	87.98	89.10
10	50.20	69.10	57.76	58.30
11	65.00	85.90	73.36	75.20
12	67.90	88.40	76.10	71.30
13	85.40	96.00	89.64	90.20
14	69.00	69.50	69.20	71.00
15	74.70	81.20	77.30	77.30
16	91.60	85.10	89.00	88.40
17	74.30	87.30	79.50	80.20
18	69.60	85.80	76.08	75.00
19	59.80	75.10	65.92	66.80
20	61.30	81.60	69.42	68.60
21	73.00	78.60	75.24	75.60
22	68.20	88.90	76.48	78.60
23	56.10	74.60	63.50	64.00
24	72.20	88.50	78.72	75.60
25	82.20	74.70	79.20	75.00
26	86.10	77.30	82.58	83.40
27	58.80	72.20	64.16	62.50
28	56.30	58.20	57.06	64.20
29	73.50	85.70	78.38	76.40
Avg.	69.48	79.89	73.64	73.23

The hypothesis of the experiment is that the average values of the final assessment (F) based on the boundary control according to Formula (32) correlate with the average values of the estimates obtained on the basis of the model under study (M). In the experiment, the hypothesis will also be tested that there is a correlation between the mean values of the boundary control (Att) and the model estimates (M).

The results of the experiment will make it possible to determine the quality of the proposed multicriteria model, which is not worse than the totality of teacher evaluations (60%) and computer testing (40%).

However, the proposed multicriteria model for computer testing, which takes into account the number of answers, the level of questions complexity depending on the time spent for answer, the time of the entire test, doubts, the number of omissions, and the use of additional programs, allows us to bring the % of successful random answers to the minimum, thereby increasing the objectivity of the level of knowledge assessment, having made them close to an assessment of the teacher, but not of the computer program.

During the experiment, the following data were obtained (Table 3).

3.2. Experiment Processing. The null hypothesis H_0 is a statistical hypothesis to be verified by statistical data, the results of observations included in the sample. Of the possible statistical hypotheses, choose the one that assumes equity, and it is the most important for further conclusions.

An alternative hypothesis is a statistical hypothesis that is considered valid if the null hypothesis is incorrect.

Statistical criterion is a rule according to which on the bases of observation results, a decision is made whether to accept or reject the null hypothesis. In this article, the experimental test of the model adequacy is chosen as the criterion or the rule.

The article puts forward the null hypothesis that the coefficient of doubt, introduced into the model as one of the factors, affects the objective assessment of the knowledge of the person being tested.

In this regard, for a more objective assessment of test results, a multicriteria knowledge assessment model is proposed. This model was compiled and tested in previous works of the authors.

This article experimentally proves the effect of the additionally introduced factor, which indicates the level of doubt of the person being tested on an effective assessment of his knowledge. This shows the adequacy of the proposed model. In the article, this type of system evaluates according to the following criteria:

- (1) Time spent to respond
- (2) Number of correct answers
- (3) Time of the entire test
- (4) Level of doubt
- (5) Skipping questions
- (6) Using the additional functions

One of the most difficult tasks is to take into account consideration of doubts depending on the complexity of the issue. In order to keep records of doubts depending on the level of question complexity, it is necessary that the program be with a preliminary choice of the answer, and subsequently with its confirmation.

Consequently, it all comes down to the number of preliminary answers, i.e., to the number of doubts. In this regard, the authors came to the conclusion that the coefficient of doubt level included in the model depends on the following variables:

- (i) Number of missed operations
- (ii) Amount of unconfirmed information
- (iii) Reaction to decision-making
- (iv) Level of difficulty
- (v) Number of logical chain interrupts
- (vi) Doubtful user actions

The multicriterial model is valid only for adaptive testing, which takes into account the number of answers, the level of complexity of questions depending on the time spent on their answer, the time of the entire test, doubts, the number of omissions, and the use of additional functions.

According to the experimental data, the multicriteria model of knowledge testing described in the article objectively assesses the level of knowledge of the person being tested and sets the appropriate assessment.

As a result of processing the obtained data, the Pearson correlation coefficients were calculated:

$$K_{xy} = \frac{\sum(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \cdot (y_i - \bar{y})^2}} \quad (33)$$

where x_i refers to values of the variable X , y_i refers to values of the variable Y , \bar{x} is the average value for variable X , and \bar{y} is the average value for variable Y .

Figure 3 shows the relationship between the values of the intermediate control (Att) and the examination evaluation of computer testing (Exm).

The correlation coefficient between Att and Exm = 0.65 (K_1).

The figure shows that the results of the assessments delivered by the teacher (Att) are weakly correlated with the assessments of the computer testing exam.

Figure 4 shows the relationship between the values of boundary control (Att) and computer testing based on the multicriteria model (M).

The correlation coefficient between Att and M = 0.94 (K_2).

Figure 5 shows the relationship between the final grade for the course (F) and computer testing based on the multicriteria model (M).

The correlation coefficient between F and M = 0.95 (K_3).

4. Discussion

The hypotheses that have been advanced are confirmed by the results of processing the experimental data, which show that the model evaluation (based on the multicriteria approach) and the final assessment, as well as the model and assessment of the boundary control obtained on the basis of standard approaches, have a high level of coincidence, which is confirmed by a high degree of pair correlation of K_2 , K_3 , and visual confirmation (Figures 4 and 5).

Thus, it can be argued that there is a sufficiently high functional connection (close to linear) between the final assessment for the course and computer testing on the basis of the multicriteria model, which makes it possible to exclude the boundary control conducted directly by the instructor, without violating the high degree of objectivity in assessing student knowledge. This result is of great importance for the organization of distance learning with the automated assessment process.

5. Related Work

Tojo et al. [3] aim to implement the prelearning and group discussion in a consistent e-learning system, which develops the human-robot interaction for the purpose of creating a social, autonomous robot capable of conversing using various human-like methods such as body language, hand gestures, facial expressions, gaze, and touching.

Joksimovic et al. [4] put forward a model, which intended to guide the future work studying the association between contextual factors, student engagement, and learning outcomes.

According to Firat et al. [5], motivation that initiates and sustains behavior is one of the most significant components of learning in any environment.

Farhan et al. concluded that students' interaction and collaboration using the Internet of Things (IoT) based on interoperable infrastructure is an effective learning practice [6]. Measuring student attention is an essential part of educational assessment.

Eom and Ashill argue that a significant reduction in dependent and independent variables and their measures are necessary for building an e-learning success model, and such

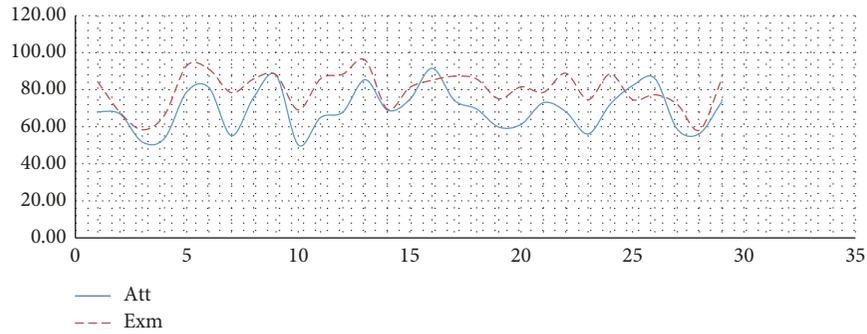


FIGURE 3: The relationship between the values of the boundary control (Att) and the examination evaluation of computer testing (Exm).

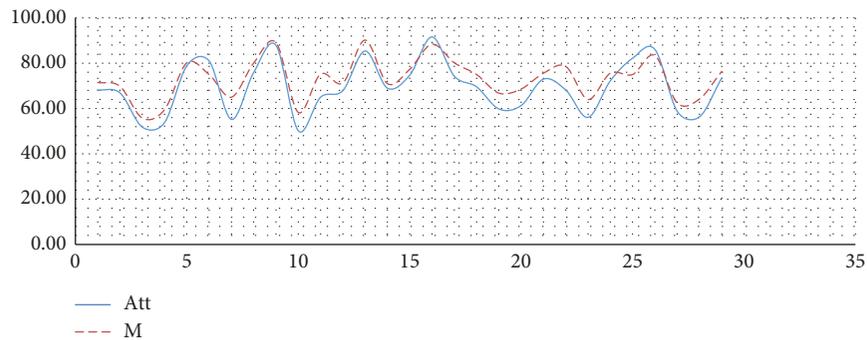


FIGURE 4: The relationship between the values of the boundary control (Att) and computer testing based on the multicriteria model (M).

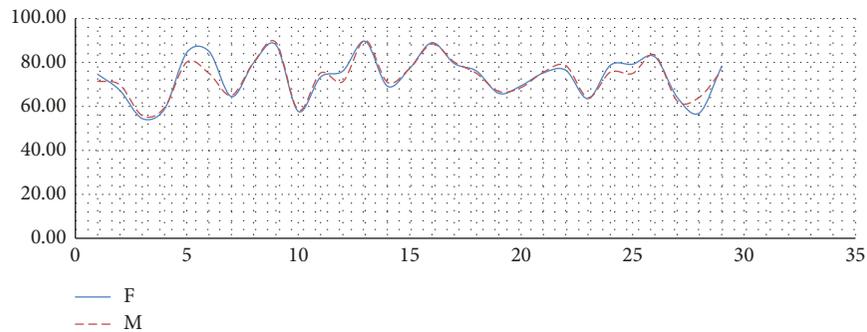


FIGURE 5: The relationship between the final grade for the course (F) and computer testing on the basis of the multicriteria model (M).

a model should incorporate the interdependent (not independent) nature of e-learning success [7].

Bhattacharya et al. present an intelligent recognizer of the cognitive state of an e-learner as an integral part of the confidence-based e-learning (CBEL) system [8]. It addresses the problem of providing technology-driven pedagogical support to an e-learner to achieve the desired cognitive state of mastery, which is endowed by high levels of both knowledge and confidence.

Marcelino et al. offered Learning Computational Thinking concepts which have gained an importance in the last years [9].

Debiec mentioned that, in this paper, a case study is presented based on one offering of an introductory digital system course taught with a combination of learner-centered strategies selected to overcome these barriers and improve student’s performance [10].

Gul et al. identified various issues and challenges faced by massive open online courses (MOOCs) while offering open online courses to vast number of learners [11].

Farid et al. propose a sustainable quality assessment approach (model) for the e-learning systems keeping software perspective under consideration [12].

Bradac and Walek present the proposal, design, and implementation of a new approach to adaptive e-learning systems [13]. Adaptivity is considered as an ability of the system to adapt to student’s knowledge and characteristics.

Ren et al. present an online course applicability assessment (OCAA) to assist learners in course selection [14]. Three main characteristics are taken into consideration, including “learning style,” “learning behavioral type,” and “prior knowledge” which considerably affect the e-learning effectiveness.

Cohen and Anat focus on the fact that identifying the changes in student activity during the course period could help in detecting at-risk learners in real time, before they actually drop out from the course [15].

In recent years, designing adaptive e-learning systems has become one of the striking topics of discussion in the literature. Hamada and Hassan describe a case study on the integration of the learning style index into an adaptive and intelligent e-learning system [16].

Levina et al. studied the continuous information development of all spheres of education: integration of new knowledge, accessibility of information technologies and computer facility aids, professionalization, and computerization of educational activities [17].

Sweta and Lal proposed learner model mines and learner's navigational access data and found the learner's behavioral patterns which individualize each learner and provide personalization in the learning process according to their learning styles [18].

Bralic and Divjak blended a learning model where a MOOC has been integrated in a traditional classroom [19]. The learning outcomes-based approach was implemented that supported a balanced student workload.

Gregori et al. aim to design and analyze the implementation of a number of guidelines that allow us to effectively unify a high-quality teaching methodology and the use of new technologies in distance learning [20].

Quintana and Perez have analyzed massive, open, and online model course, describing their advantages and disadvantages, besides the formative and motivating factors that guide the students, not only to enroll in particular courses but also to request for their own space to be able to create a MOOC [21].

For the first time, a multicriteria method for measuring the level of knowledge of students during computer testing based on the level of doubt for E&M learning has been described by the authors in a number of works, as discussed by Serbin et al., [22–24] and applied for distance training.

6. Conclusion

The multicriteria model allows to objectively assess the student's knowledge level and, accordingly, to estimate the evaluation during computer testing. The model includes an original method of measuring the level of doubt, as a function of the time, and complexity of issues differentiated by thematic blocks. In addition, the assessment model itself being multicriteria, besides the function of doubt, includes the number of correct answers, the testing time, the use of additional programs, spent time on each response, the number of periodic passes in the testing process, and more. In diagnosing the level of knowledge based on computer testing, use of this method will reduce the probability of the influence of an accidental correct answer ("guessing") on the final result and to obtain the most objective assessment.

The proposed model includes the following factors:

(i) Number of answers

- (ii) Level of complexity of the questions depending on the time
- (iii) Spent on their response
- (iv) Time of all testing
- (v) Doubt
- (vi) Number of passes
- (vii) Use of additional programs

The practical significance of this model lies in the objectivity of assessing the level of knowledge, where computer testing is applied, especially relevant for the organization of distance learning using E&M learning. Additional effects of the application of this model may be the reduction of the time for organization, intermediate control procedures, labor costs, operating costs, processing and analysis costs, and others, research planned in subsequent publications.

In conclusion, we can make a statement that the null hypothesis put forward by the authors is proven experimentally and the study demonstrates the original methodology for measuring the level of doubt of the tested students in the information and training system. The use of this methodology allows making the most objective decision on the assessment of students' knowledge.

In diagnosing the level of knowledge based on computer testing, the use of this methodology makes it possible to reduce the probability of the random factor influence, such as "guessing" on the final result and to obtain the most objective assessment. Moreover, to control the learning process, measuring the level of doubt allows taking into account the psychological characteristics of human behavior during training.

Data Availability

The raw data that support the findings of this study are available from the corresponding author upon reasonable request.

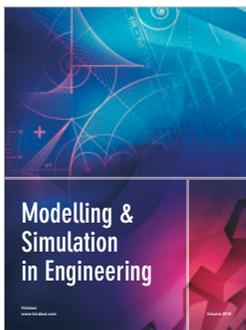
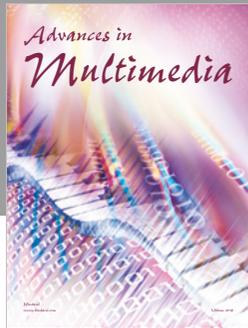
Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] M. West and S. Vosloo, *UNESCO Policy Guidelines for Mobile Learning*, 2015, <https://unesdoc.unesco.org/ark:/48223/pf0000219641>.
- [2] G. Vinogradov and V. Kuznetsov, "Modeling the agent's behavior with subjective notions of the choice situation," *Artificial Intelligence and Decision-Making*, vol. 3, pp. 58–72, 2011.
- [3] T. Tojo, O. Ono, N. B. M. Noh, and R. Yusof, "Interactive tutor robot for collaborative e-learning system," *Electrical Engineering in Japan*, vol. 203, no. 3, pp. 22–29, 2018.
- [4] S. Joksimovic, O. Poquet, V. Kovanovic et al., "How do we model learning at scale? A systematic review of research on MOOCs," *Review of Educational Research*, vol. 88, no. 1, pp. 43–86, 2018.

- [5] M. Firat, H. Kilinc, and T. V. Yuzer, "Level of intrinsic motivation of distance education students in e-learning environments," *Journal of Computer Assisted Learning*, vol. 34, no. 1, pp. 63–70, 2018.
- [6] M. Farhan, S. Jabbar, M. Aslam et al., "IoT-based students interaction framework using attention-scoring assessment in e-learning," *Future Generation Computer Systems*, vol. 79, no. 3, pp. 909–919, 2018.
- [7] S. B. Eom and N. J. Ashill, "A system's view of e-learning success model," *Decision Sciences Journal of Innovative Education*, vol. 16, no. 1, pp. 42–76, 2018.
- [8] S. Bhattacharya, S. Roy, and S. Chowdhury, "A neural network-based intelligent cognitive state recognizer for confidence-based e-learning system," *Neural Computing and Applications*, vol. 29, no. 1, pp. 205–219, 2018.
- [9] M. J. Marcelino, T. Pessoa, C. Vieira, T. Salvador, and A. J. Mendes, "Learning computational thinking and scratch at distance," *Computers in Human Behavior*, vol. 80, pp. 470–477, 2018.
- [10] P. Debiec, "Effective learner-centered approach for teaching an introductory digital systems course," *IEEE Transactions on Education*, vol. 61, no. 1, pp. 38–45, 2018.
- [11] S. Gul, I. Mahajan, H. Shafiq, M. Shafi, and T. A. Shah, "Massive open online courses: hype and hope," *DESIDOC Journal of Library & Information Technology*, vol. 38, no. 1, pp. 63–66, 2018.
- [12] S. Farid, R. Ahmad, M. Alam, A. Akbar, and V. Chang, "A sustainable quality assessment model for the information delivery in e-learning systems," *Information Discovery and Delivery*, vol. 46, no. 1, pp. 1–25, 2018.
- [13] V. Bradac and B. Walek, "A comprehensive adaptive system for e-learning of foreign languages," *Expert Systems with Applications*, vol. 90, pp. 414–426, 2017.
- [14] Y. Ren, Z.-X. Dai, X.-H. Zhao, M.-M. Fei, and W.-T. Gan, "Exploring an on-line course applicability assessment to assist learners in course selection and learning effectiveness improving in e-learning," *Learning and Individual Differences*, vol. 60, pp. 56–62, 2017.
- [15] A. Cohen, "Analysis of student activity in web-supported courses as a tool for predicting dropout," *Educational Technology Research and Development*, vol. 65, no. 5, pp. 1285–1304, 2017.
- [16] M. Hamada and M. Hassan, "An enhanced learning style index: implementation and integration into an intelligent and adaptive e-learning system," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 8, pp. 4449–4470, 2017.
- [17] E. Levina, A. Masalimova, N. Kryukova et al., "Structure and content of e-learning information environment based on geo-information technologies," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 8, pp. 5019–5031, 2017.
- [18] S. Sweta and K. Lal, "Personalized adaptive learner model in E-learning system using FCM and fuzzy inference system," *International Journal of Fuzzy Systems*, vol. 19, no. 4, pp. 1249–1260, 2017.
- [19] A. Bralic and B. Divjak, "Integrating MOOCs in traditionally taught courses: achieving learning outcomes with blended learning," *International Journal of Educational Technology in Higher Education*, vol. 15, pp. 1–16, 2018.
- [20] P. Gregori, V. Martínez, and J. J. Moyano-Fernández, "Basic actions to reduce dropout rates in distance learning," *Evaluation and Program Planning*, vol. 66, pp. 48–52, 2018.
- [21] J. G. Quintana and J. M. Perez, "The students empowerment in the sMOOC," *Revista Complutense de Educacion*, vol. 29, no. 1, pp. 43–60, 2018.
- [22] V. V. Serbin, A. Syrymbayeva, and K. Tolebayeva, "Multi-criteria decision-making model for information learning system: a critique of the level of doubt," *International Journal of eGovernance and Networks*, vol. 3, no. 1, pp. 56–65, 2015.
- [23] V. Serbin and Y. Gorbunov, "Analysis of MPP-systems stress testing based on big data," *International Journal of Advances in Electronics and Computer Science*, vol. 4, no. 2, pp. 56–60, 2017.
- [24] V. V. Serbin and A. M. Syrymbayeva, "Research of multi-criterial decision-making model for educational information systems," *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, vol. 16, no. 5, pp. 946–951, 2016.



Hindawi

Submit your manuscripts at
www.hindawi.com

