



Copyright © 2017 by Academic Publishing
House Researcher s.r.o.

All rights reserved.

Published in the Slovak Republic

European Journal of Contemporary Education

ISSN 2304-9650

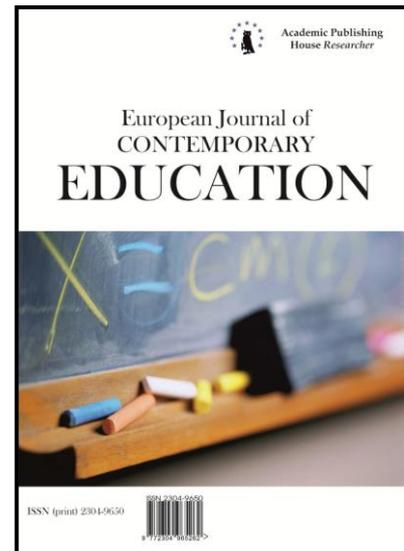
E-ISSN 2305-6746

2017, 6(2): 254-263

DOI: 10.13187/ejced.2017.2.254

www.ejournal1.com

WARNING! Article copyright. Copying, reproduction, distribution, republication (in whole or in part), or otherwise commercial use of the violation of the author(s) rights will be pursued on the basis of Russian and international legislation. Using the hyperlinks to the article is not considered a violation of copyright.



Testing Methodology in the Student Learning Process

Tatiana N. Gorbunova ^{a, b, *}

^a Moscow State University of Civil Engineering (National research university MGSU),
Russian Federation

^b Moscow Polytechnic University, Russian Federation

Abstract

The subject of the research is to build methodologies to evaluate the student knowledge by testing. The author points to the importance of feedback about the mastering level in the learning process. Testing is considered as a tool. The object of the study is to create the test system models for defence practice problems. Special attention is paid to the reliability of simulated tests, and to their differentiating ability to assess knowledge. The author pays significant attention to the learning aspect of tests, that assumes the student's choice of the method to solve proposed test tasks. Various methods are suggested, including a model of fuzzy estimation of knowledge based on testing results, which allows to evaluate not only the result obtained, but also the solution method.

Open and multiple choice problems, offered to students and evaluated by the binary system, are chosen as a methodology for creating tests. An algorithm to test the hypothesis of normal distribution using the Shapiro-Wilk criterion is proposed. The scoring scales and the creating tests methodology are also described.

There are the following conclusions. The first, the possibility of using tests to evaluate students' understanding of the acquired material. The second, the possibility of including testing in the educational process. The technology for estimating knowledge using a fuzzy model is proposed, which is in good agreement with the methodology of training itself. The main result is the developed technology of testing with automatic option development and result processing. The proposed testing algorithm can be included in the general methodology of studying the discipline.

* Corresponding author

E-mail addresses: tngorbunova@yandex.ru (T.N. Gorbunova)

Keywords: testing, normal distribution, Shapiro-Wilk criterion, evaluation scale, the assimilation criterion, open test, multiple choice test, correlation, statistical processing, reliability rating, membership function, linguistic scale, fuzzy set.

1. Introduction

The learning process is a complex system. A key factor of the consistency of its functioning is the possibility of obtaining and deciphering the feedback needed for all participants in the educational process. The teacher, receiving the necessary information about the level of student achievements, can adjust the educational process for its optimization. The trainee also gets the opportunity to get self-control and self-diagnosis of his training.

Today, there are many ways to set up such feedback, including in the form of various tests, that, when performing a full set of measures, can be served as a relatively objective tool of pedagogical diagnostics to organize an effective feedback system (Baayen, 2008; Bartram, 1995; Kenneth, 2012; Reeve, 2003; Fisher, 2014).

Pedagogical testing serves not only the purposes of monitoring. As V. Avanesov notes (Avanesov, 1989) one of the functions of pedagogical testing is the teaching function, which most clearly manifests itself in programmed learning. The purpose of this study was to develop a testing technology and include it in the learning process.

2. Preparation and testing

MGSU Department of Applied Mathematics has been testing students in mathematics for several years. The positive experience of using the developed method described in (Safina et al., 2015; Osipov et al., 2016) made it possible to extend this method to Computer Science. From the entire course of this discipline, a semester about the study of the basics of programming in the MATLAB environment, and the introduction into numerical methods using knowledge of linear algebra, was selected. Therefore, this semester can be considered as the application of mathematical models in Computer Science, and this module can be considered as interdisciplinary.

This course of Computer Science consists of a set of lectures and 7 lab tasks. Student should take a differentiated test to pass this course. Topics are studied:

1. The solution of the system of linear algebraic equations (SLAE) by the Gauss method;
2. The solution of SLAE by iterative methods, such as the simple iteration method and the Seidel method;
3. calculation of the inverse matrix by the Gauss method;
4. Calculation of eigenvalues and eigenvectors of the matrix using the power method;
5. Numerical integration by methods of rectangles, trapezoids and Simpson;
6. Solution of the nonlinear equation by half-division method and the Newton method;
7. The least square method for constructing an optimal line.

The implementation of each lab test consists of three stages:

1. perform a manual calculation to study the functioning of the mathematical model and obtain a test result;
2. implement the computer program, in this case in the MATLAB environment, and compare the results with the manual calculation;
3. defence the lab test.

The last stage of lab test defence is applied as a test. This semester ends with a differentiated test that evaluates the student knowledge using four-point grade system.

Therefore, the task that was posed to the author is whether it is possible to develop a testing system that could be used to grade students by the results of the test.

Testing was conducted among the first and second year students of the Moscow State University of Civil Engineering in the discipline of Computer Science.

3. Test material preparation

All technology of tests modeling should be defined based on the requirements. In this case, as such a goal, the problem was formulated to assess student's understanding of mathematical models and numerical methods of the class of required problems. To achieve this goal, the following technologies were used to develop the tests, which help to streamline and efficiently organize the

test based on the theory of develop tests by V.S. Avanesov (Avanesov, 2015; Maiorov, 2001). Here are briefly the main statements that have been used.

4. Rules of the tasks preparation

1. Each test topic was assigned to 4 test tasks. The formulation of all tasks was written in a unified logical form of the statement in the form of an affirmative sentence and in a laconic form, excluding the wrong interpretation.

2. Each group of questions contained two questions (theoretical and computational) of multiple choice type, when the student chooses the correct or several correct answers from the proposed options, and two open type calculation tasks, where the student had to solve the problem and write down the result.

3. Using the multiple choice and open form of the task allows to build tasks with increasing complexity and, thus, to increase tasks learning ability.

As an example, below are four test tasks on the topic of iterative methods for solving SLAE.

1. Chose systems with diagonal predominance

$$\bullet \begin{cases} 3x - 2y + z = 1 \\ 3x - 11y + 7z = 2 \\ x + 2y - 3z = 3 \end{cases} \bullet \begin{cases} 2x_1 + 3x_2 - x_3 = 9 \\ x_1 - 2x_2 + x_3 = 3 \\ x_1 + 2x_3 = 2 \end{cases} \bullet \begin{cases} 2x_1 + 10x_2 - 3x_3 = 38 \\ -3x_1 - 12x_2 + 13x_3 = -82 \\ x_1 + 3x_2 - 5x_3 = 27 \end{cases}$$

$$\bullet \begin{cases} 4x_1 + x_2 - 3x_3 = -12 \\ -x_1 - 12x_2 + 11x_3 = -28 \\ x_1 + 3x_2 - 5x_3 = 51 \end{cases}$$

2. Solve the system of equations $\begin{cases} 5x_1 + 2x_2 - x_3 = -1 \\ 3x_1 + 10x_2 + x_3 = 9 \\ x_1 + x_2 - 5x_3 = -2 \end{cases}$ by the simple iteration method.

Complete 1 step. Set $x_1^0 = 0, x_2^0 = 0, x_3^0 = 0$ as an initial approximation. Write x_2^1 as a result.

3. Let there be given a system of linear equations $\begin{cases} x_1 + 5x_2 = 1 \\ 2x_1 + 2x_2 = 3 \end{cases}$. It must be written in the

following form for the convergence of the simple iteration method

- $\begin{cases} x_2 = (1 - x_1)/5 \\ x_1 = (3 - 2x_2)/2 \end{cases}$
- $\begin{cases} x_1 = 1 - 5x_2 \\ x_2 = (3 - 2x_1)/2 \end{cases}$
- $\begin{cases} x_1 = 2x_1 + 5x_2 - 1 \\ x_2 = 2x_1 + 3x_2 - 3 \end{cases}$
- $\begin{cases} x_2 = x_1 + 6x_2 - 1 \\ x_1 = 3x_1 + 2x_2 - 3 \end{cases}$

4. Using the simple iteration method (3 iterations), determine the 1st column of the inverse matrix to the matrix A: $A = \begin{bmatrix} 6 & -15 \\ 15 & 8 \end{bmatrix}$

The last task allows you to check the understanding of not only the iterative process, but also the understanding of the method of calculating the inverse matrix. Thus, it allows the student to perform the analysis and synthesis actions to obtain the solution of the problem, which corresponds to the upper levels of the Bloom taxonomy (Kim, 2007).

4. The application of the principle of facetedness allowed to compile 30 similar tasks for each of the 28 types of tasks. The test variants were randomly generated using the original technique described in (Safina et al., 2015; Osipov et al., 2016). "The author's program in the form of a macro in Visual Basic generates 30 parallel individual test cases, distributing tasks between subjects in a random way, which ensures the uniqueness of the set of assignments for each student in all study groups"

5. The test is given to students in paper form. The method of solving and the result of the performance of students is recorded in written form.

The main advantage of this method, according to the authors, is the methodological organization of the learning process itself. Testing is part of this process. The defence of lab test should show an understanding of the methods of solving a certain class of problems. All computational tasks were formulated in such a way that they can be solved both manually (Gorbunova et al., 2015), and on a computer in the MATLAB system or with the help of other software tools, for example, Excel. Therefore, it was the paper version of the test that enabled the student to make a choice in favor of one or another method of solution, by writing the solution.

Of course, the shortcomings of this form of test are also obvious. The check is performed manually by the teacher using a file with answers that is generated by the program at the same time as individual task variants. Thus, the verification is fast enough.

And one more undoubted advantage of this form is the possibility of individual analysis of the test in the presence of the student "in the hot pursuit", as well as generalized analysis for the whole group.

6. Testing was limited by 60 minutes.

Task # 1. Evaluation of test quality

The results of the testing were statistically processed using a table processor Microsoft Excel (Shtainer, 2000; Gorbunova, Zhuravleva, 2014). Each task was evaluated on a two-point scale: 0 – the problem was solved incorrectly, 1 – the problem was solved correctly. According to the classical theory of tests, the test result of the test subject is determined by a test score - the sum of the points of each solved test.

5. Evaluation the hypothesis of the normal distribution of test scores

The testing was conducted for students in three groups of the first year and two groups of the second year, with a total of 98 people. The results of the statistical analysis were as follows.

It is known that for a normative-oriented test designed to rank test subjects per their level of knowledge using standardization methods, the distribution curve of the test scores should be symmetrical and close to the Gaussian curve.

There are many ways to check the correspondence of the distribution of points to normal. The choice was made in favor of the method described in (Popov, 2012; Gorbunova, 2017) for small sample groups. The method consisted of several stages: the build of histograms of the distribution of individual scores for groups of subjects depending on the number of solved problems, the analysis of statistical characteristics such as: median, mode, error of the mean, etc. The final decision was made after the Shapiro-Wilk criterion was fulfilled. This criterion is applied just for a small set <50.

The algorithm for calculating the Shapiro-Wilk criterion is sufficiently technological and can be implemented by Excel, or as an independent application. The method was described in (Zalyazhnykh, 2014).

It was concluded that the distribution of test scores in all groups is normal based on study results.

6. Correlation between tasks

Average correlation between topics in [table 1](#).

Table 1. Average correlation between topics

	topic 1	topic 2	topic 3	topic 4	topic 5	topic 6	topic 7
A1	0,324	0,291	0,292	0,129	0,310	0,215	0,237
B1	0,337	0,238	0,321	0,314	0,161	0,350	0,295
C1	0,286	0,157	0,225	0,351	0,239	0,176	0,244
A2	0,271	0,168	0,241	0,297	0,182	0,161	0,254
B2	0,282	0,198	0,281	0,164	0,173	0,289	0,275

According to the recommendations of VS Avanesov (Avanesov, 2005; Kim, 2007) the correlation between topics should not be too high (≤ 0.3), otherwise the topics begin to duplicate each other.

The average correlation of the results by assignment and individual scores are shown in table 2.

Table 2. The average correlation of the results by assignment and individual scores

	topic 1	topic 2	topic 3	topic 4	topic 5	topic 6	topic 7
A1	0,688	0,660	0,545	0,448	0,572	0,491	0,504
B1	0,567	0,467	0,771	0,730	0,492	0,631	0,523
C1	0,539	0,487	0,514	0,705	0,463	0,278	0,517
A2	0,673	0,710	0,651	0,725	0,561	0,492	0,671
B2	0,597	0,680	0,678	0,692	0,547	0,548	0,586

This result is also correlated with Avanesov’s recommendations. The conclusion made that text assignments for different learning groups are consistent with the classical theory of test development and can be used in the learning process.

Task # 2. Evaluate the degree of mastering the material using the test.

The ratio of expectation value to standard deviation is used to estimate of differentiating ability of the test namely the ability to divide students by various levels based on the test score. The differentiating ability of the test is considered satisfactory if the ratio is about 3 or more (Avanesov, 1989).

Table 3. The sufficiently high differentiating ability of the test

	\bar{X} / s_x
Group A1	2,97
Group B1	3,12
Group C1	3,71
Group A2	4,89
Group B2	5,58

The data in the table 3 indicates a sufficiently high differentiating ability of the test. Several methods were used to achieve this goal

7. Rating scale application

According to the goal, namely, the lab test defence in the test version, each topic is considered separately in the assessment. The result of each topic is the score of the same four-point scale: “excellent”, “good”, “satisfactory” and “bad” (“very bad” and “bad” merged into one rating). The topic is considered mastered and appropriately passed if three or more problems are solved. Considering this situation, the result of the test was adjusted, namely the resulting score was calculated considering the submitted topics. The results of this approach were consistent with the normal distribution of 28 points for a four-point scale.

The transfer points from one scale to a five-point scale (1, 2, 3, 4 and 5) was carried out according to the formula

$$O_n = S_5 \left(\frac{p - m_{28}}{S_{28}} \right) + m_5 \tag{1}$$

where O_n – a normalized rating in the final five-point scale; p -points from 0 to 28 the original selection; m_{28} – the mathematical expectation of the studied population; S_{28} - the standard deviation of the study sample; S_5 и m_5 – respectively the standard deviation and the mathematical expectation of the final scale.

Table 4. Assessment scales

Scale of assessment by reference points		Correction of the scale considering the passed topics		The normal distribution scale, O_n	
Initial 'raw' scores	Score by test	Points	Score by test	Points	Grade
1	2	3	4	5	6
0–12	2	0–13	2	0–11	2
13–17	3	14–18	3	12–16	3
18–22	4	19–22	4	17–22	4
23–28	5	23–28	5	23–28	5

According to the above results, the adjusted range of low scores shifts from the higher score. The range of good grades is also narrowed, which is consistent with the recommendations from the paper (Kim, 2007; Dubas, 1990).

The result of using this technique was the opportunity to obtain an appraisal judgment about the level of students' knowledge, as well as identify topics that are learned and not mastered by students and their subsequent study. At the final exam, students received assignments and questions on topics with the lowest scores.

The described technique allowed to make a conclusion about the understanding of the topic and get an estimate on the four-point scale system. The topics where students got low scores were proposed to pass by the traditional way of oral communication with the teacher, and additional problems were given to students to solve.

As a result, students got the final grade by one or two points higher on average than the test results. Undoubtedly, the additional preparation of students on topics affected here. On average only 5–7 % of students got a final grade on the differential test coinciding with the test.

8. Understanding level estimate for a group.

The next question raised whether it was possible to consider the individual characteristics of the group's training during the assessment of the test, which the teacher considers when working with the group? Groups study by different curricula in different subjects and have some common characteristics.

The criterion of mastering j -th topic q_j is the ratio of the number of students who solved three or more problems from j -th topic, to the total number of students. Accordingly, a criterion of not understanding (level of complexity) of the j -th topic is a number equal to $p_j = 1 - q_j$ which is the ratio of the number of students solved less than three problems from the topic to the total number of students.

These criteria for all 7 topics were determined by the test results separately for each group of students. The average value \bar{p} of the problem simplicity in the group is equal to the ratio of the average number of completed tasks \bar{X} to the test volume $n = 28$; mean complexity $\bar{q} = 1 - \bar{p}$. The results are presented in the [table 5](#).

Table 5. The complexity level criteria of topics for groups \bar{q}

group	topic 1	topic 2	topic 3	topic 4	topic 5	topic 6	topic 7
A1	0,364	0,636	0,682	0,955	0,514	0,591	0,545
B1	0,245	0,418	0,455	0,864	0,691	0,418	0,482
C1	0,211	0,500	0,611	0,722	0,722	0,611	0,667
A2	0,193	0,512	0,594	0,681	0,545	0,623	0,692
B2	0,201	0,598	0,498	0,647	0,613	0,681	0,593

As can be seen from the table, topics for diverse groups represent a different level of difficulty.

Therefore, it was suggested to consider the level of complexity for topic understanding. The following scale was introduced.

Table 6. Assessment scale

Difficulty level	Additional points
[0;0,3]	-1
(0,3;0,7]	0
(0,7 ;0,9]	+1
(0,9;1)	+1

Thus, we have received a mechanism for regulating the border whether a topic has been mastered or not for a given group.

The usage of this technique allowed to obtain estimates for the test already at 40 % coinciding with the final grade.

9. Multitasking model of fuzzy estimation of knowledge

Previous technique is an intermediate option between the classical version and innovative approaches. It can be adjusted not only by the rating scale, but the assessment of the task itself cannot be as 'solved/do not solved' method. The other way is incomplete and not very precise methods of solving problems selected by the student. It is also necessary to consider the way of constructing the test itself, using problems of different difficulty levels. And, since students receive test assignments in paper form and write the solution also on paper, it becomes possible in a more complex scale of assessment.

A model of fuzzy estimation of knowledge is offered, that allows to obtain a multivalued quantitative evaluation of the decisions made on their qualitative descriptions. With respect to the task formulation on the learning topics, the tester can determine the degree of truth for each answer by constructing the so-called function of its belonging to the truth estimation scale.

Thus, there is a fundamental opportunity to formulate and present to the examiner the tasks where the task is evaluated corresponding to a multivalued linguistic scale of the following type:

$$L = [\text{'right'}, \text{'incomplete'}, \text{'inaccurate'}, \text{'undefined'}, \text{'wrong'}]$$

As the basic technique was taken the method described in the paper ([Scherbina, Smyikova, 2011](#)).

The correctness of each variant of the answer is characterized by the membership function specified on the linguistic variable, as the base set of discrete values of the linguistic evaluation scale is used for. The set of answers of each test task is represented by a fuzzy set, where each element is a pair of values ('answer variant' for the multiple-choice test task / 'performance level' for the open, 'membership function'). The degree of "total" correctness of the learner's answers to

all presented test tasks is estimated during testing. This indicator is calculated using the fuzzy algebra apparatus, by constructing the membership function of the set of selected answers applied to the linguistic scale (Urobotov, 2011).

The procedure for setting the degree of truth to the proposed answers for each test task is determined

$$(V, M_E, L) \rightarrow A_j = \{(a_{ij} ; \mu_{ij})\}, \quad i = \overline{1, M}, j = \overline{1, N} \quad (2)$$

where V is the set of test tasks; M_E – master model of student knowledge; L – linguistic variable that determines the scale of evaluating the correctness of answers;

$$A_j = \{(a_{ij} ; \mu_{ij})\}, \quad i = \overline{1, M}, j = \overline{1, N}$$

fuzzy set of possible answers (here M is the number of variants of answers to the test task $v_j \in V$ or level of this job, N – number of test

tasks (in our case - 28); a_{ij} – i-th the answer to the j-th task, μ_{ij} – the membership function that

determines the degree of truth of the answer a_{ij}).

The student selects the solution of multiple choice problem and shows completion and method to solve the open problem for each j-th job. Based on actual responses $\tilde{a}_{ij} \in A_j$ and the

corresponding membership function $\tilde{\mu}_{ij}$ the correct population is calculated as a membership

function μ_Σ normalized with respect to the number of tasks in the test

$$\mu_\Sigma = \frac{1}{N} \sum_{j=1}^N \tilde{\mu}_{ij} \quad (3)$$

To determine the final score, a comparison is made between the obtained membership function and the reference function.

Table 7. Reference functions of the membership of the final estimates μ_t paper (Scherbina, Smyikova, 2011)

Evaluation O_t	‘right’	‘incomplete’	‘inaccurate’	‘undefined’	‘wrong’
Unsatisfactory	0	0	0,1	0,3	1
Satisfactory	0,2	0,4	0,9	0,7	0,3
Good	0,7	0,9	0,7	0,3	0,1
Excellent	1	0,3	0,1	0	0

As the final evaluation, the value of O_t is taken, for which the Hamming distance between the corresponding standard membership function μ_T (table 7) and the obtained membership function

of the set of selected student responses μ_Σ (3) is minimal

$$\rho(T; \Sigma) = \sum_{t=1}^T |\mu_T(x_t) - \mu_\Sigma(x_t)|$$

$$O = \{O_t, \min \rho(T; \Sigma)\}$$

As an example, we give the following.

Here is the obtained membership function for a student from group A1, who answered 28 questions and did not answer earlier:

$$\tilde{\mu}_{ij} = [18 / \text{'correct'}, 6 / \text{'incomplete'}, 2 / \text{'inaccurate'}, 1 / \text{'undefined'}, 1 / \text{'incorrect'}]$$

Table 8. Final grade/Hamming distance

final grade	Hamming distance
Unsatisfactory	2.12
Satisfactorily	2.4
Good	1.72
Excellent	0.54

The minimum distance is 0.54, so the student can get “Excellent” grade. This student could get no more than “Good” grade according to the first technique. The teacher receives a complete picture about the assimilation of the student.

10. Conclusion

The main result of this work is the development of a testing technique to pass Computer Science lab test, which is presented as an algorithm. Its steps include: the task formulation, testing, statistical processing, testing the hypothesis of a normal distribution, obtaining estimates scales, and analysis of topic understanding.

The proposed multitasking model of fuzzy assessment of knowledge allows you to flexibly assess the level of knowledge and understanding of the student. This increases the level of the training function of the test. The model itself allows to use many linguistic variables for different problem sets. For example, for open and multiple choice problems.

All steps are implemented automatically using Excel macros, and MATLAB functions. As a further development, it is intended to fully computerize the entire process without losing its methodological merit to distance learning.

References

- [Avanesov, 1989](#) – *Avanesov V.S.* (1989). *Osnovy nauchnoi organizatsii pedagogicheskogo kontrolya v vysshei shkole* [Fundamentals of scientific organization of pedagogical control in higher education]. Moscow: MISIS, 1989. 168 p.
- [Avanesov, 2005](#) – *Avanesov V.S.* (2005). *Forma testovykh zadaniy* [Form of test tasks]. Moscow: Tsentr testirovaniya. 156 p.
- [Baayen et al., 2008](#) – *Baayen R.H., Davidson D.J., Bates D.M.* (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008. Vol. 59, is. 4. pp. 390–412.
- [Bartram, 1995](#) – *Bartram D.* (1995). The development of standards for the use of psychological tests in occupational settings: The competence approach. *The Psychologist*, 5. pp. 219-223.
- [Dubas, 1990](#) – *Dubas V.* (1990). Ob otsenivanii znaniy pri programmirovannom kontrole [On the evaluation of knowledge in programmed control]. *Fizika v shkole*, №3. p. 83.
- [Fisher, 2014](#) – *Fisher W.P., Jr.* (2014). The Central Theoretical Problem of the Social Sciences. *Rasch Measurement Transactions*. Vol. 28: 2. pp. 1464-1466.
- [Gorbunova, Zhuravleva, 2014](#) – *Gorbunova T.N., Zhuravleva T.Yu.* (2014). *Avtomatizirovannyi laboratornyi praktikum po informatike Osvoenie raboty v MS Excel 2007* [Automated laboratory practical work on computer science Mastering work in MS Excel 2007]. Resource document Monografiya. Vuzovskoe obrazovanie, EBS «IPRbooks», Saratov, 2014. 77 p. <http://www.iprbookshop.ru/20699> Accessed 06.12.2016.
- [Gorbunova et al., 2015](#) – *Gorbunova T.N., Chalmov I.D., Chernykh S.I.* (2015). Ruchnoi i avtomatizirovannyi podkhod pri izuchenii informatsionnykh tekhnologii [Manual and automated approach to the study of information technology]. *Nauka, obrazovanie, obshchestvo: aktual'nye voprosy i perspektivy razvitiya: sb. nauch. tr. M.* p. 22-23.

Gorbunova, 2017 – *Gorbunova T.N.* (2017). Tehnologiya testirovaniya v protsesse izucheniya informatiki [Testing technology in the process of studying computer science]. *Pedagogika i prosveschenie*. № 1. pp. 74-85. DOI: 10.7256/2454-0676.2017.1.22125. URL: http://e-notabene.ru/pped/article_22125.html

Kenneth, 2012 – *Kenneth D.R.* (2012). Measuring Liberal. Conservative Voting Tendencies among U.S. Senators. *Rasch Measurement Transactions*. Vol. 26: 2. pp. 1366-1367.

Kim, 2007 – *Kim V.S.* (2007). Testirovanie uchebnykh dostizhenii [Testing of educational achievements]. Ussuriisk: UGPI. 214 p.

Maiorov, 2001 – *Maiorov A.N.* (2001). Teoriya i praktika sozdaniya testov dlya sistemy obrazovaniya [Theory and practice of creation of tests for the education system]. M.: «Intellect-tsentr», 2001. 296 p.

Osipov et al., 2016 – *Osipov Yu.V., Safina G.L., Vetukhnovskii F.Ya.* (2016). Modelirovanie testov po matematike [Simulation tests on math]. *Otkrytoe i distantsionnoe obrazovanie*. № 3 (63). pp. 69-77.

Popov, 2012 – *Popov O.A.* (2012). Psikhologicheskaya statistika [Psychological statistics]. [Elektronnyi resurs]: Resource document: [http:// www.psystat.at.ru](http://www.psystat.at.ru) Accessed 20.12.2016

Reeve, 2003 – *Reeve B.B.* (2003). Item response theory modeling in heart outcomes measurement. *Expert Review of Pharmacoeconomics and Outcomes Research*. 3(2). pp. 131-145.

Safina et al., 2015 – *Safina G.L., Osipov Yu.V., Kerimova D.Kh., Krasovskaya I.A.* (2015). Poluavtomaticheskaya sistema testirovaniya po matematike [Semi-automatic testing system for mathematics]. *Otkrytoe i distantsionnoe obrazovanie*. № 2 (58). pp. 56-62.

Scherbina, Smyikova, 2011 – *Scherbina Yu.V., Smyikova K.V.* (2011). Mnogoznachnaya model nechetkogo otsenivaniya znaniy po rezultatam testirovaniya [Multivalued model of fuzzy estimation of knowledge based on test results]. *Vestnik MGUP imeni Ivana Fedorova. Izdatelstvo Moskovskiy Gos. Uni.-t pechati im. Ivana Fedorova*, pp. 220-226.

Shtainer, 2000 – *Shtainer G.* (2000). Visual Basic 6.0 dlya prilozhenii [Visual Basic 6.0 for applications]. Moscow: Laboratoriya Bazovykh Znanii, 2000. 832 p.

Uhobotov, 2011 – *Uhobotov V.I.* (2011). Izbrannyye glavyye teorii nechetkikh mnozhestv [Selected chapters of the theory of fuzzy sets]. Chelyabinsk: Izd-vo Chelyab. gos. un-ta, 2011. 245 p.

Zalyazhnykh, 2014 – *Zalyazhnykh V. V.* (2014). Statisticheskie raschety pri planirovanii i obrabotke rezul'tatov ispytaniy [Statistical calculations in the planning and processing of test results]. Severnyi (Arkticheskii) federal'nyi un-t im. M. V. Lomonosova. Arkhangel'sk: SAFU. 82 p.