



Increasing Learning Efficiency Using Adaptive Testing Technology

Yuldashev Abdusamat Halimovich

Senior Teacher, Andijan Mechanical Engineering Institute, Andijan, Uzbekistan

Ermatov Axror Baxtiyorjon Ogli

Assistant, Andijan Mechanical Engineering Institute, Andijan, Uzbekistan

Copyright: Original content from this work may be used under the terms of the creative commons attributes 4.0 licence.

ABSTRACT

Purpose: The article describes a set of software developed for adaptive testing technology in the implementation of an objective assessment of students' knowledge. There is also information about the possibility of computerizing education, reducing the unproductive live work of teachers, preserving the methodological potential of experienced professors, installing computer software for management.

Methods: It is noted that the experiments were carried out by 2nd year students of the Andijan Machine-Building Institute in the direction of "Ground transport systems and their operation."

Results: The research results are presented in the form of some data by means of mathematical statistical processing. The Pearson, Kolmagorov and Romanovsky criteria were also used to check the accuracy of the study results.

Conclusion: it is stated that a software package aimed at creating a technology for remote and adaptive testing without the participation of the human factor will allow processing the results of experiments in educational and research centers and achieving them in practice.

KEYWORDS

Adaptive test, database, standard deviation, arithmetic mean, probability.

INTRODUCTION

In the context of the rapid development of technologies and information processes in the modern world, special attention is paid to assessing the scientific potential of higher

educational institutions and research centers, the development of information systems for monitoring information and communication technologies, as well as targeted research.

“One of the key indicators in the ranking of higher educational institutions is the need for students to use residual knowledge and their future use, the use of new technologies and methods aimed at the development of competitive specialists, especially specialties and special disciplines that are difficult to master. In this regard, for an objective assessment of the scientific potential of universities and research centers, the development of test items with guaranteed quality indicators, as well as the development of modules, methods, algorithms and software systems that allow its systematic application.

SOLUTION TECHNIQUE

Thus, to solve the above problems, we propose the purposeful use of a computer program in the PHP programming language [3]. At the same time, the operator has the ability to track each student's grades and test assignments. The test subjects create a login and password to enter the program and transfer the current and final control. Based on this login and password, students can complete test tasks by entering the program. The advantage of using the software is that it has a test case database and a MySQL database. There are also options for downloading them via MC Word or MS Excel to add them to the database. The program

allows testing by selecting a database from existing test items. As a result, the reliability and validity of the test items will increase based on the invariant test items. This may be the only effective way to objectively assess students' knowledge.

The quality of the test items (reliability and validity) is controlled by a computer program. At the same time, tests are assessed on the basis of a certain set of parameters for test items, which are clearly answered during the control control [4-15]:

$$r_{bis} = \frac{M_T - M_{HT}}{S_x} \cdot \sqrt{p_j q_j} \quad (1)$$

Where: MT is the arithmetic mean of students who answered the task correctly; MNT is the arithmetic mean of students who react incorrectly to the task; Sx - entire test, standard deviation for all tests; p_jq_j - Standard deviation for task j.

If the majority of the subjects answer exactly one question during the test, the program evaluates this question as a low level of difficulty, that is, a simple question. Thus, the computer program automatically determines the degree of difficulty of the question. When determining the level of knowledge, the program being tested differs from the test tasks in terms of the level of complexity (Fig.1.)

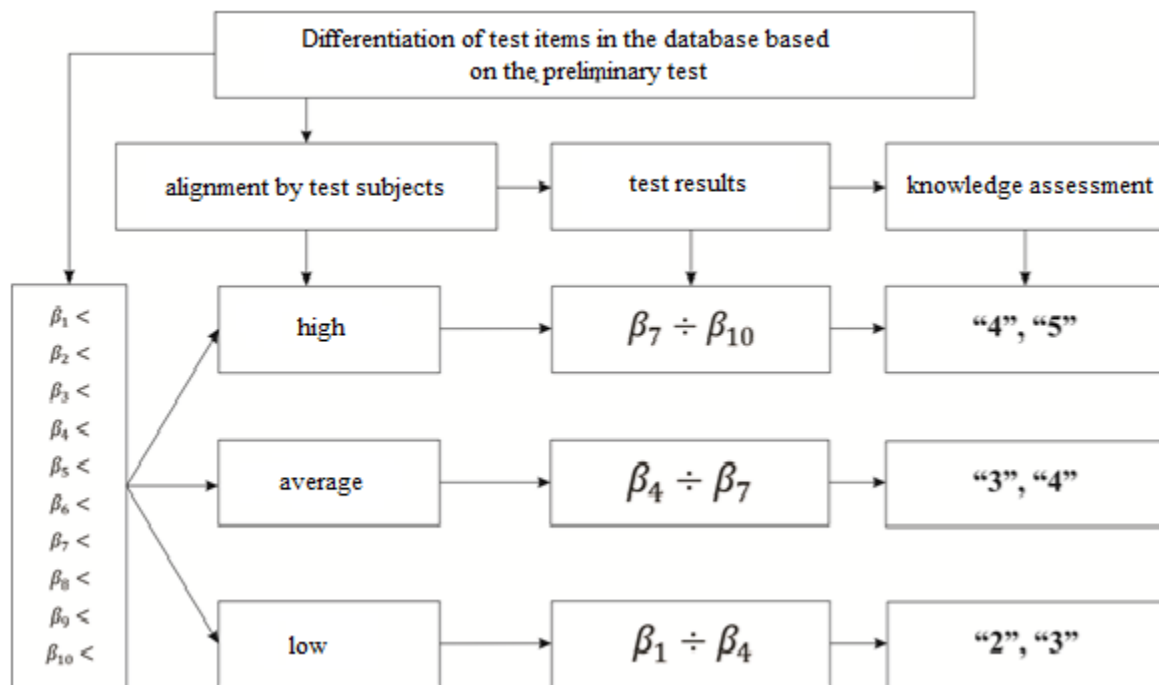


Figure 1. Technology for evaluating test items by degree of complexity.

The program also takes into account the likelihood that students will accidentally find the correct answer on test items. It is advisable to use the three-parameter mathematical model of Birnbaum. Birnbaum's

three-parameter mathematical model, which provides a quality assurance icon for invariant test items, can be used as the analytical basis of the method [4-15]:

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

Where: P_i is the probability that the subject will correctly answer the task i -test,
 a_i - a feature of the test,
 θ - the level of knowledge of the subject,
 b_i - the complexity of the test item.
 c_i is the probability that the subject will accidentally find the correct answer ($c_i = 0.25$ for a test with four answers, $c_i = 0.2$ for five and 0.1 for ten).

Formula (2) serves as a key analytical expression for improving the quality of invariant test items.

In this case of pyramids, such a structure is convenient, since the complexity of the student's knowledge and test items or the

complexity of the test shows the stages of difficult access to the student's knowledge. It also determines the rating of the assessment.

The computer program provides one step-by-step test with an average level of difficulty in each of the 150 selected test items. If the student answers the test items correctly, the

pyramid will go up the stairs, otherwise it will go down. As a result, they will have an objective and fair assessment of their knowledge (Fig. 2. and Fig. 3.).

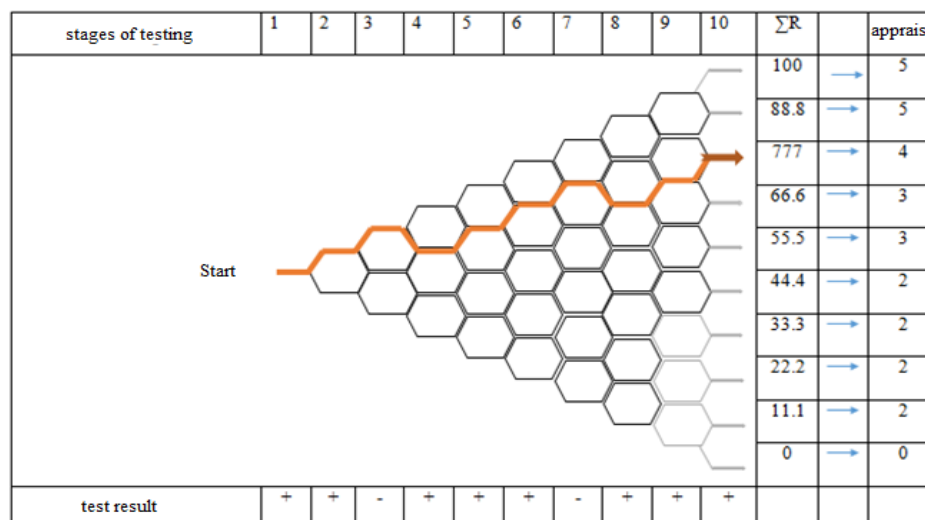


Figure: 2. Technology of adaptive testing to assess the knowledge of students and listeners

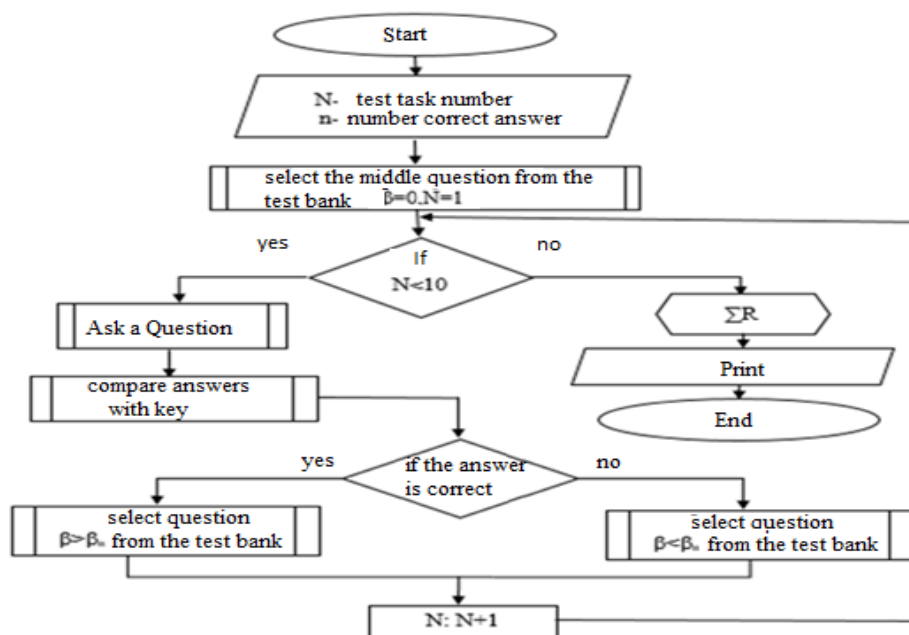


Figure: 3. Block diagram of the adaptive testing technology algorithm

The advantage of this method and technology is that it equates to a fair assessment of students' knowledge, saving a lot of labor and energy [16–19].

The only drawback is that the process must be conducted with the participation of experienced trainers and senior operators.

The aforementioned software packages are designed to develop creative thinking along with the intended use of invariant test items, increase their enthusiasm through an objective assessment of students' knowledge, as well as much more than "56", and "70", "75" and "80" ... to achieve the best results.

THE RESULT OF THE STUDY

As a result of experiments, observations and gathering information about the phenomenon

12	9	17	19	16	13	17	18	20
16	16	14	19	11	17	20	11	
1	25	17	14	12	22	19	9	

This random set of numbers must be put in a certain order. To do this, first of all, it is necessary to determine the smallest and largest values of the evaluation points. In this example, this is 0 point and 30 point. The difference between these values is called the range of the variation series $R = 30$ points. [15]

The next processing operation is to divide the variation series into intervals. Usually almost the number of intervals is chosen 5-12, but you can determine the number of intervals by an approximate formula.

Where N is the sample size, i.e. the number of available values of the trait.

of interest to us, statistical material accumulates, i.e. A set of values for one or more parameters, usually in a chaotic state. This is the so-called unordered variation series of values. Information was gathered about the students' knowledge, and such an unordered variation series of values was obtained by their examinations by statements. Total data obtained for 25 students:

In our example, $K = 6$. Since the span of $R = 30$, the width of the interval $h = 3.75$ It should be noted, however, that the above formula for the number of intervals does not always give a successful number. Sometimes he uses the formula or even selects the number of intervals without any calculation within the above limits.

From the set of values of attributes, their number is calculated, which fall into each interval, putting in the corresponding line of table 1. The strokes must be grouped into five units so that it is more convenient, then counted. This frequency table should have six columns:

Table 1.

№ interval	Border spacing	Mid interval	Frequency		Frequency
			Strokes	Number	
1	2	3	4	5	6
1	0-3,75	1,875		0	0,00
2	3,75-7,5	5,625		0	0,00
3	7,5-11,25	9,375	III	4	0,13
4	11,25-15	13,125	IIII	5	0,22
5	15-18,75	16,875	IIIIII	8	0,34
6	18,75-22,5	20,625	IIIIII	7	0,24
7	22,5-26,25	24,375	I	1	0,07
8	26,25-30	25,125		0	0,00

Further processing of the variation series consists in calculating the average value of the attribute or reliability index and the dispersion characteristic around this average value, i.e. standard deviation or standard values.

Having the data in Table 1, you can build a histogram of the distribution of the studied reliability index, for example, as in the example above, the reliability of the tests. For

this, the values of the exponent and the boundaries of the intervals of the variational series are plotted on the abscissa axis, and the frequency m_{ei} or frequency in absolute values or in percentages is plotted along the ordinate axis. Rectangles are built on each interval, the height of which is equal to the frequency or frequency in this interval.

The histogram for the given example of the variation series is built in Fig.4. apparently, it has a fairly regular symmetric shape [8,9].

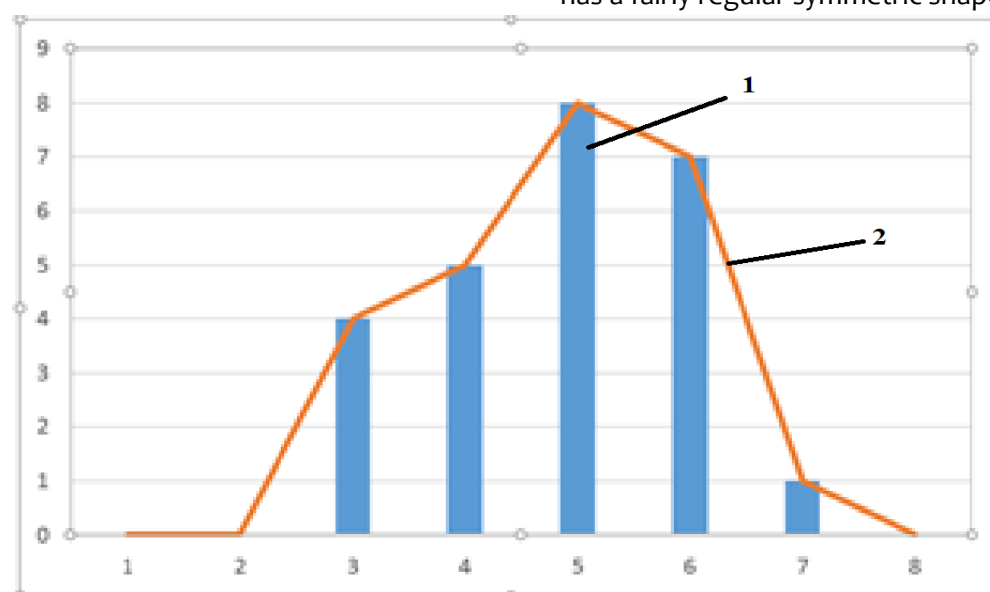


Figure 4. The histogram, the distribution of the more experienced frequency (1) and a flattening curve of the theoretical curve (2).

The obtained histogram is approximated by a curve of the theoretical distribution law of a random variable. This is necessary to

accomplish in order to obtain a more general distribution, which could be extended to the entire set of objects studied. In addition, the

creative laws of distribution are well studied, so they can be used to more fully characterize the studied reliability indicators.

The most common in the theory of reliability are the law of normal distribution, the Weibull law, the exponential law, and some others. The choice of the most appropriate theoretical distribution law is made by the appearance of the histogram and the magnitude of the coefficient of variation. Close to symmetric histograms and when most likely approximated by a normal law. Constantly decreasing histograms with $V = 1$ indicate an indicative law. The Weibull distribution law is usually best suited for asymmetric histograms at $V = 0.5 \dots 0.7$ [16].

When choosing a theoretical distribution law, it is necessary to take into account the physical nature of the phenomenon being studied, the observation of which gave a variation series of values of the indicator.

If observations were made of gradual object failures caused by a series of independent causes and the influence of many random

factors, then the normal distribution law should be used; distribution of failures due to aging of materials are in good agreement with the Weibull law. The combined effects of aging and wear lead to the distribution of failures close to the log-normal law. Distributions of sudden failures are in good agreement with the indicative law.

To verify the correctness of the choice of the theoretical distribution law, it is necessary to calculate the so-called compliance criteria, which indicate the degree of discrepancy between the theoretical and experimental frequencies or frequencies. For each distribution law considered further, specific instructions will be given on determining the criteria for compliance.

Data processing according to the normal distribution law

The appearance of the histogram in the example of section 1 shows that it can be approximated by a normal distribution curve, the formula for which is as follows:

$$f(L) = \frac{N * \Delta L}{S_L \sqrt{2\pi}} e^{-\frac{(L-\bar{L})^2}{2S_L^2}}$$

To calculate the parameters of the normal distribution and build its theoretical curve, table 2 is compiled.

Definitions of the sample mean i. average student knowledge:

$$\bar{L} = \frac{\sum_{i=1}^k L_i m_{\theta i}}{N} = 16.29$$

Since $N > 30$, the calculation of the standard deviation is carried out according to the uncorrected formula:

$$S_L = \sqrt{\frac{\sum_{i=1}^k L_i^2 m_{\theta i}}{N} - \bar{L}^2} = 4.39$$

Knowing the mean value and standard deviation for each interval, we determine the value of the parameter and fill in the column of the 7th table 2.

Table 2.

№ interval	Mid interval L_{pi}	Experimental frequency m_{pi}	L_i^* m_{pi}	L_{2i}^* m_{pi}	$L_i - \bar{L}$	Para-meter $t = \frac{L_i - \bar{L}}{S_i}$	Normalized function $f(t)$	Theoretical		Accumulated frequencies		Frequency difference
								Frequency, m_{Ti}	Frequency, $m_{\pi Ti}$	Experimental	Theoretical	
1	2	3	4	5	6	7	8	9	10	11	12	13
1	1,875	0	0	0	-14,40	-3,429	0,001125	0	0,025115	0	0	0
2	5,625	0	0	0	-10,65	-2,536	0,016128	0	0,360000	0	0	0
3	9,375	4	37,5	351,562	-6,90	-1,643	0,104166	1	2,325146	4	1	3
4	13,125	5	65,625	861,328	-3,15	-0,750	0,303151	1	6,766761	9	2	7
5	16,875	8	135	2278,125	0,60	0,143	0,397532	1	8,873492	8	3	5
6	20,625	7	144,375	2977,734	4,35	1,036	0,234893	1	5,243139	17	4	13
7	24,375	1	24,375	594,141	8,10	1,929	0,062539	0	1,395955	18	4	14

According to the obtained values of t in the table of the normalized function $f(t)$, we determine its values (entered in column 8).

Now we determine the constant coefficient in the density formula for the normal distribution, equal to:

$$\frac{N * \Delta L}{S_L} = \frac{25 * 3.75}{4,2} = 22,32143$$

where ΔL is the width of the interval of the variation series.

We multiply the values of the normalized function (column 8) by the value of this coefficient and put the result in column 9, rounding it to an integer - these will be the theoretical frequencies m_{Ti} , and divided by their sum, i.e. $\sum m_{Ti}$ will have theoretical frequencies in column 10.

Having in table 2 the values of the experimental and theoretical frequencies of the normal distribution of operating time, it is possible to calculate the agreement criteria for checking the correctness of the choice of the theoretical distribution that approximates the empirical histogram.

Consider the three criteria of consent.

a) Pearson Consensus Criteria [17,18] or χ^2 . This criterion represents the sum of the squares of the deviations of the theoretical frequencies from those experienced in each interval, referred to the theoretical frequency in this interval.

$$\chi^2 = \sum_{i=1}^k \frac{(m_{pi} - m_{Ti})^2}{m_{Ti}}$$

When calculating according to this formula, it should be borne in mind that the absolute frequency in the interval should be at least five, therefore, with its smaller values, adjacent intervals should be combined. In our example, the first and last intervals have an absolute frequency equal to three, so these intervals are combined: the first in the second and sixth with the seventh. Thus, in the calculation is taken $K=4$

According to the table, the probability $P(\chi^2)$ of coincidence of the experimental and theoretical frequencies is determined, which should be as close as possible to unity. To determine it, it is necessary to know the number of degrees of freedom, which is equal to:

$$\chi^2 = \sum_{i=1}^k \frac{(m_{oi} - m_{Ti})^2}{m_{Ti}} = \frac{(0-0)^2}{0} + \frac{(0-0)^2}{0} + \frac{(4-2)^2}{2} + \frac{(5-7)^2}{7} + \frac{(8-9)^2}{9} + \frac{(7-5)^2}{5} + \frac{(1-1)^2}{1} = 3,48254$$

According to the proposed formula, the probability values

$$P(\chi^2) = 0,31$$

It is known [10,11] that for $P(\chi^2) < 0,3$, i.e. at 30% probability of discrepancy is considered significant and the law of normal distribution cannot be used

Thus, in our example, we obtained a condition $P(\chi^2) > 0,30$ satisfying that the discrepancy between the theoretical and experimental frequencies is insignificant and we can accept the curve of the law of normal distribution for smoothing experienced histograms.

b) The consent criteria of Professor Romanovsky [17,18]. For its calculation, the ratio

$$a = \frac{|\chi^2 - \tau|}{\sqrt{2\tau}}$$

is determined, where τ - is the number of degrees of freedom.

If the indicated ratio in absolute value is less than two, then the discrepancy between the theoretical and experimental frequencies can be considered insignificant and a normal distribution law can be adopted. For the applied Romanovsky criterion is

$$a = \frac{|3,48254 - 2|}{\sqrt{2 \cdot 2}} = 0,74$$

those. less critical ($0,74 < 2$)

c) The consent criteria of Academician Kolmogorov [17,18] establish the closeness of theoretical frequencies by an experimental method of comparing their integral distribution functions. The calculation of the criterion λ is carried out according to the formula

$$\lambda = \frac{D}{\sqrt{N}}$$

where D is the maximum value of the modulus of the difference of the accumulated theoretical experimental frequencies, N is the sample size (the number of observations)

in our example, the value $D = 3$ and therefore, the criterion λ is equal to:

$$\lambda = \frac{3}{\sqrt{25}} = 0,6$$

In the probability table $P(\lambda)$, we find the corresponding value. $P(\lambda) = 0,9983$, which means that the discrepancies between the theoretical and experimental frequencies are random and the curve of the law of normal distribution approximates the experimental histogram well.

CONCLUSION

In order to improve the quality of education and residual knowledge, students are

encouraged to develop independent learning, a creative approach to learning, computer technology to ensure the objectivity of the assessment of knowledge, as well as the

development of specific test items and their intended use in teaching basic subjects.

Thus, we can conclude that the advantages and differences of adaptive testing technology are that it provides an objective and objective assessment of the student's knowledge, as well as low cost and ease of testing.

REFERENCES

1. А.Х.Юлдашев, Н.А.Алиева, З.З.Джумабаева. Агроинженерия фанларини ўқитишда талабалар билимини ҳолисона баҳолаш ва тест топшириқларининг сифат кўрсаткичларини кафолатлаш. // Ўзбекистон аграр хабарномаси. - Тошкент. 2016. - №3. Б. 109-113.
2. Джумабаев А.Б., Юлдашев А.Х., Джумабаева З.З. Узлуксиз таълим тизимида ўқув жараёнини мазмун ва сифат жиҳатдан баҳолашда техник тартибга солиш масалалари. // STANDART илмий-техника журнали. - Тошкент. 2015. - №4. - Б. 34-36.
3. №DGU05086, 05.03.2018 “Тест топшириқларини қўллаш билан талабалар билимининг сифат кўрсаткичларини кафолатлаш учун дастурий мажмуа”, // А.Б.Джумабаев,А.Х.Юлдашев, И.Юсупов,Н.Н.Рашидов.
4. Birnbaum A. Some Latent Trait Models and Their Use in Statistical Theories of Mental Test Scores. – Reading Mass.: Addison-Wesley, 1978. – 568p.
5. Ингенкамп К. Педагогическая диагностика / пер. с нем. – М.: Педагогика, 1991. – 240 с.
6. Д. Халперн. Психология критического мышления. Санкт-Петербург: Питер. 2000, 503 с.
7. Rasch G. Probabilistic model for some intelligence and attainment tests. Chicago. Univ. of Chicago Press. 1980, 199 p.
8. Аванесов В.С. Научные проблемы тестового контроля знаний: монография. – М.: Исследовательский центр проблем качества подготовки специалистов, 1994. – 135 с.
9. Аванесов В.С. Композиция тестовых заданий: учебник для преподавателей вузов, техникумов, училищ, учителей школ, гимназий и лицеев. –М.: Центр тестирования, 2002. – 239 с.
10. Аванесов, В.С. Математические модели педагогического измерения: научное издание / В. С. Аванесов // Педагогические измерения. -№4. - 2008. - С. 3-24.
11. Анастаси, А. Психологическое тестирование: пер. с англ. / А. Анастаси, С. Урбина. - 7-е междунар. изд. - М.; СПб.: Питер, 2006. - 688 с.
12. Талызина Н. Ф. Теоретические основы контроля в учебном процессе. -М.: Знание, 1983. -96 с.
13. Батурин, НА. Универсальный интеллектуальный тест УИТ СПЧ-М. Руководство. Челябинск-Санкт-Петербург / НА. Батурин, НА. Курганский. - Челябинск., 2003. - 56 с.
14. Беспалько В. П. Педагогика и прогрессивные технологии обучения. -М., 1995.-. 41-62.
15. Беспалько В.П. Основы теории педагогических систем: (Проблемы и методы психолого-педагогического обеспечения технических обучающихся систем). – Воронеж: Изд-во Воронежского ун-та, 1997. – 304 с.
16. Альтиментова Д.Ю., Гданский Н.И. Адаптивные модели компьютерного обучения // Педагогическая информатика. – 2015. – № 2. – С. 83–92. 210
17. Чельшкова М.Б. Теория и методология адаптивного тестирования в образовании. - М.: Исслед. центр, 2001. 165 с.

-
18. Computerized Adaptive Testing: From Inquiry to Operation /Ed. by W. A. Sands, B. K. Walters, I. R. McBride. - ASC, 1997.- 292 p.
 19. Handbook of Modern Item Response Theory /Ed. By Wim J. van der Linden and Ronald K. Hambleton. ASC, Univ. Ave, 1997 - 510 p.