UDC 001.891.572

N. Jaremchyk, A. Kravchenko

RESEARCHES OF ADAPTIVE TESTING ALGORITHMS IN LEARNING SYSTEMS

Introduction

Now there is considerable development of information technologies in education and it needs new modern approach to the construction of the learning systems. Before they were a set of hypertext electronic materials and tests, but now it is the systems with intellectualization of many functions. The basic requirements to such systems are known: intellectuality, openness, flexibility and adaptiveness at the organization of process of teaching and control. The adaptiveness of control process solves by using adaptive testing. The adaptive testing exists for a long time [1]. But works of a choice and optimization of algorithms of the adaptive testing are continuing. The main purpose of algorithms development is to get advantages before the ordinary (unadapted) testing. Basic traditions, which have already made at the adaptive testing are using of the stratified tests, questions of which choose from the test space (bank of questions) in, and this test space is distributed on the difficulty levels; analysis of results of the current testing, after which next question will be chosen from a top level if the answer is right, and from lower – if the answer is wrong.

Mathematical processing of results gives a quantitative estimation of difficult questions – logit of difficulty – like as $\delta_i = \ln\left(q_i / p_i \right)$, where q_i – part of the wrong answers, p_i – part of the right answers to the question i from set of tested student, and estimation of knowledge – logit of knowledge – $\theta_j = \ln\left(p_j / q_j \right)$, where p_j – part of the right answers of tested student j, q_j - part of the wrong answers of tested student j to all questions.

Problem definition

Researches of algorithms adaptive testing are executed by means of modeling testing procedure and processing of modeling's results. The model of a student's condition from the point view of student's learning curve, model of the stratified test space, model of student's reaction on a certain question from the test space, model of the tutor, that is to say model of actions for level definition of a feedback are used for that purpose.

Besides, it was conducted a preliminary choice of investigated testing kind and strategy. On adaptation parameter it is possible to allocate the next kinds of testing: with adaptation based on the contents of questions, with adaptation based on difficulty of questions and with adaptation by quantity of questions. There is the testing with adaptation based on difficulty of questions was researched in this work. In turn, on testing strategy the testing with adaptation based on difficulty of questions distinguishes to pyramidal and “flex level” testing [2]. At pyramidal testing procedure of testing begins with a question of the middle difficulty level, at "flex level" testing – with the difficulty level, which chooses student. It is necessary to notice that "flex level" testing on the strategy causes the big difference according to quantity of tested questions and theirs dependency on a student’s self-estimation. That is to say, if the student correctly estimates the possibilities for a correct estimation level of his knowledge it is necessary the minimum quantity of questions. And the largest quantity of questions will be at start from the lowest difficulty level for the student whose preparation corresponds to the top level and contrary. Since one of research problems is choice of quantity test’s questions that pyramidal testing is chosen for research and modeling.

System of adaptive testing

The block diagram of adaptive testing system is presented on fig. 1.

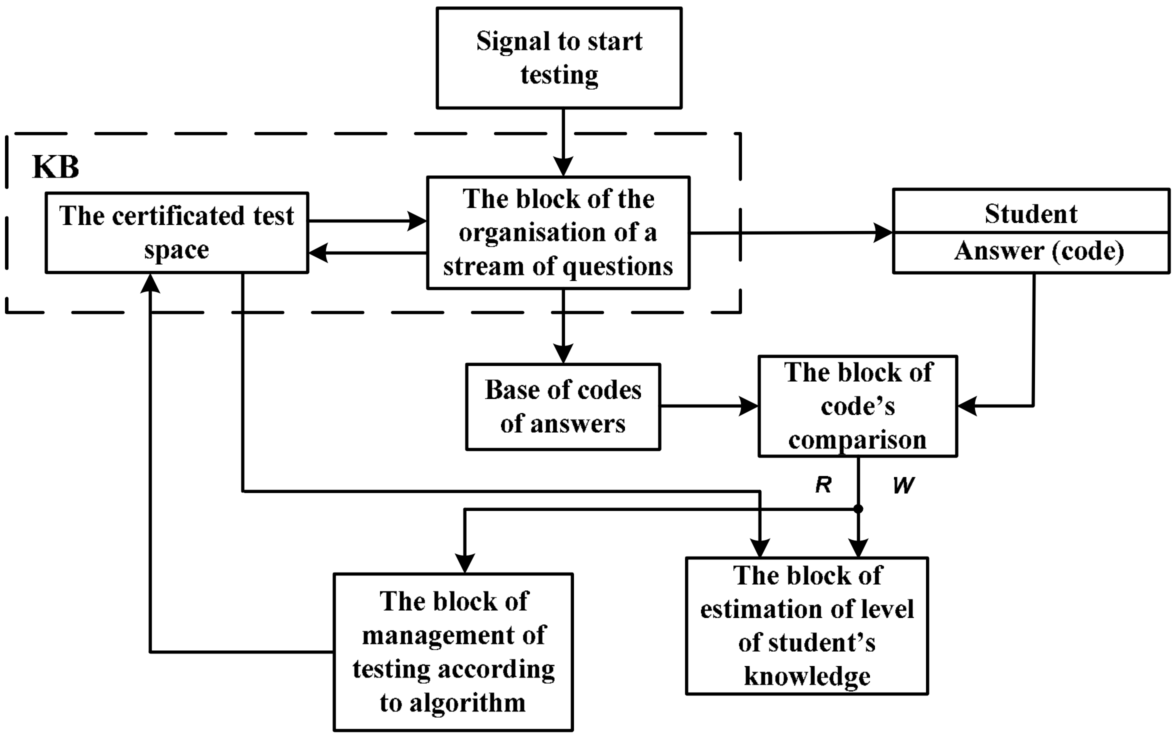


Fig. 1. The block diagram of adaptive testing system

On the figure 1 **KB** is the knowledge base.

If there is a signal to start testing, the question will be given to a student from the middle difficulty level. According to the number of questions from base answer's code, corresponding code actuates and compares to code of student's answers. The right answer sets off and the corresponding signal comes to the block of management testing process which operates according to the accepted testing algorithm. At a following step difficulty of a question increases or remains the same for one more step of testing. After finishing testing process an average difficulty of the test student j calculates with the next formula:

$$\bar{\delta}_j = \frac{1}{n} \sum_{i=1}^{n} \delta_{ij},$$

where $i = \overline{1,n}$, n – quantity of questions in the test,
and estimation of knowledge calculates like:

$$\theta = \ln\left(p_j \big/ q_j \right) + \bar{\delta}_j. \qquad (1)$$

Modeling procedure of adaptive testing

The student's model has been presented by hypothetical learning curve θ_h in logit. In the stratified test space three gradations have been allocated with average difficulty levels -3, 0, 3 logit, which are corresponding to the verbal gradation: questions are "easy", "middle" and "difficult" [3]. The Rasch model has been used as a model of the student's reaction j on a question of certain difficulty level [4]:

$$p_{ij} = \left\{ 1 + \exp\left[-(\theta_j - \delta_{ij}) \right] \right\}, \qquad (2)$$

where p_{ij} – probability that the student will answer to the question with difficulty level δ_{ij} correctly.

Making a decision about setting off the answer like a correct corresponds to ordinate a point of the characteristic function $p_{ij} = f(\delta_{ij})$, for which $\theta_j = \delta_{ij}$. That is to say, if $p_{ij} \geq 0{,}5$ at modeling it is considered that the student has answered correctly, and at $p_{ij} < 0{,}5$ – wrong.

Researching has been chosen the algorithm of adaptive testing with continuous adaptation. That is to say at the wrong answer difficulty of question went down, and at the right one difficulty of a question – raised. Results of modeling adaptive testing with quantity of questions 20 and division of questions on 3 difficulty levels are resulted in the table 1.

Modeling procedure has been realized by next steps:

Step 1. Start hypothetical learning curve is set $\theta_{hj} = 5$ logit and the middle level of difficulty questions of test space is equal $\delta_{1j} = 0$.

Step 2. Probability of the student's right answer p_{1j} to the first question will be calculated by the formula (2).

Step 3. Following accepted algorithm decision to increase the difficulty level of question is made if $p_{ij} > 0,5$.

Step 4. Probability of the student's right answer p_{2j} to the second question will be calculated.

Table 1.

Results of modeling adaptive testing with continuous adaptation
("+ " – right answer, "-" – wrong answer)

	θ_h=5 logit				**θ_h =3 logit**				**θ_h =0 logit**				**θ_h =-3 logit**			
	δ		**P_{ij}**	**P_{ij}**	**δ**		**P_{ij}**	**p**	**δ**		**P_{ij}**	**p**	**δ**		**P_{ij}**	**p**
1	0	+	0,99		0	+	0,95		0	+	0,5		0	–	0,05	
2	3	+	0,88	0,88	3	+	0,5	0,5	3	–	0,05	0,05	-3	+	0,5	0,5
3	3	+	0,88	0,78	3	–	0,5	0,25	0	+	0,5	0,5	0	–	0,05	0,05
4	3	+	0,88	0,68	0	+	0,95	0,95	3	–	0,05	0,05	-3	+	0,5	0,5
5	3	+	0,88	0,60	3	+	0,5	0,5	0	+	0,5	0,5	0	–	0,05	0,05
6	3	+	0,88	0,53	3	–	0,5	0,25	3	–	0,05	0,05	-3	+	0,5	0,5
7	3	–	0,88	0,46	0	+	0,95	0,95	0	+	0,5	0,5	0	–	0,05	0,05
8	0	+	0,99	0,99	3	+	0,5	0,5	3	–	0,05	0,05	-3	+	0,5	0,5
9	3	+	0,88	0,88	3	–	0,5	0,25	0	+	0,5	0,5	0	–	0,05	0,05
10	3	+	0,88	0,77	0	+	0,95	0,95	3	–	0,05	0,05	-3	+	0,5	0,5
11	3	+	0,88	0,68	3	+	0,5	0,5	0	+	0,5	0,5	0	–	0,05	0,05
12	3	+	0,88	0,60	3	–	0,5	0,25	3	–	0,05	0,05	-3	+	0,5	0,5
13	3	+	0,88	0,53	0	+	0,95	0,95	0	+	0,5	0,5	0	–	0,05	0,05
14	3	–	0,88	0,46	3	+	0,5	0,5	3	–	0,05	0,05	-3	+	0,5	0,5

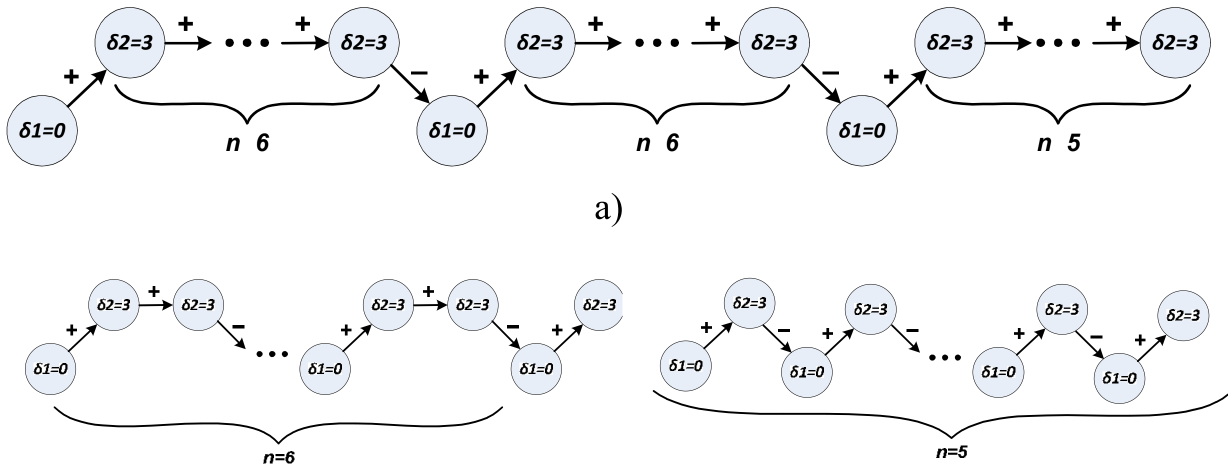
	θ_h=5 logit				θ_h =3 logit				θ_h =0 logit				θ_h =-3 logit			
	δ		P_{ij}	P_{ij}	δ		P_{ij}	p	δ		P_{ij}	p	δ		P_{ij}	p
15	0	+	0,99	0,99	3	–	0,5	0,25	0	+	0,5	0,5	0	–	0,05	0,05
16	3	+	0,88	0,88	0	+	0,95	0,95	3	–	0,05	0,05	-3	+	0,5	0,5
17	3	+	0,88	0,77	3	+	0,5	0,5	0	+	0,5	0,5	0	м	0,05	0,05
18	3	+	0,88	0,68	3	–	0,5	0,5	3	–	0,05	0,05	-3	+	0,5	0,5
19	3	+	0,88	0,60	0	+	0,95	0,95	0	+	0,5	0,5	0	–	0,05	0,05
20	3	+	0,88	0,53	3	+	0,5	0,5	3	–	0,05	0,05	-3	+	0,5	0,5

We make the decision to increase the difficulty level, but if boundary difficulty level has been reached we stay at the same level.

If difficulty level stays without changes the following meaning of probable right answers will be calculated like:

$$p_{ij} = \left(p_{ij}\right)^m .$$

If difficulty level is varies, the next meaning of probable right answer will be calculated by the formula (2). Changing process of questions difficulty level with $\theta_h = 5$ logit is presented on fig. 2.



a)

b) c)

Fig. 2. Schemes of changing difficulty level of questions with continuous adaptation for a) $\theta_h = 5$ logit; b) $\theta_h = 3$ logit; c) $\theta_h = 0$ logit

According to results of modeling it has been received meanings of θ_{je} at different quantity of questions in the test (table 2).

Table 2.

Illustration of difference between the hypothetical (set) level and empirical learning curve at the different quantity of questions in the test at testing with continuous adaptation

θ_{jh}	n=20	n=15	n=10	n=7
	θ_{je}	θ_{je}	θ_{je}	θ_{je}
5	4,8	4,27	4,6	2,98
3	2,8	2,69	2,65	1,71
0	1,5	1,5	1,5	0,88
-3	-1,5	-1,5	-1,5	-2,13

After analysis data which were received at testing modeling with continuous adaptation, it is possible to make the next conclusions. At a small quantity of questions with difficult level, results of estimation learning curve on testing θ_{je} can differ considerably from true (latent) learning curve, especially at values learning curve which are close to zero in logit (where $q_j \approx p_j$). But as researches show, the length of the test can be considerable smaller. That is, at small quantity of difficulty levels which the test space is divided into (for example, "easy", "middle" and "difficult" questions), adaptive testing can be used for a rough estimation of level at small quantity of questions.

At pyramidal testing and continuous adaptation, the quantity of questions can be equal $n = 7, 10$. At "flex level" testing, the quantity of questions should be increased approximately twice.

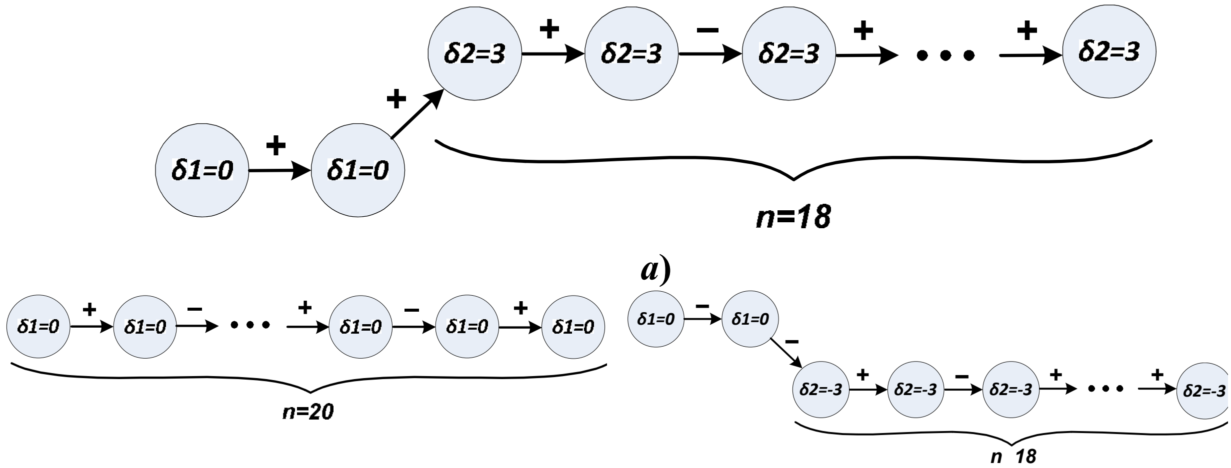
If quantity of difficulty levels which the test space is divided on is increasing, the accuracy will increase learning curve definition in testing. As researches have shown, difficulty levels should be divided non-uniformly for increasing accuracy in the middle part of a range difficulty test space. At the modeling, the next gradations a scale of difficulty levels have been established (in logit): -4; -2; -1; -0,5; 0; 0,5; 1; 2; 4. The increase in quantity of gradations of difficulty level in l times demands increase in a minimum quantity of questions in the test in $\sqrt{l}$ times. That is, the minimum quantity of questions in the test will be equal approximately 20. Results modeling with the resulted quantity of gradation for the test are presented in the table 3.

Table 3.

Results of modeling of testing with continuous adaptation

θ_{jh}	5	3	0	-3
θ_{je}	4,37	2,98	0,25	-2,98

Really, if quantity of gradations level of complexity is increased in the middle part range, the accuracy of definition of learning curve in testing will increase. Modeling with block adaptation has been done in this work. The algorithm making a decision has been complicated in the next way. The decision has been accepted on the block of questions and the quantity of questions with the same level in the block was equal two.



***a*)**

***b*)** ***c*)**

Fig. 3. Variation schemes of questions level of complexity with block adaptation ***a***) $\theta_h = 3$ logit; ***b***) $\theta_h = 0$ logit; ***c***) $\theta_h = -3$ logit

Rules making a decision of student's reaction have been presented a little bit further:

Rule 1: If two answers are correct the transition to higher level will be done.

Rule 2: If two answers are wrong, the transition to more low level will be done.

Rule 3: If one answer is wrong, and another correct, we stay at the same level of complexity.

The change process level of complexity questions with block adaptation is presented on fig. 3.

Results of modeling adaptive testing with block adaptation and quantity of 20 questions and divided questions on 3 levels of complexity are presented in the table 4.

Table 4.

Results of modeling testing with block adaptation

("+ " – right answer, "-" – wrong answer)

	θ_h =3 logit				**θ_h =0 logit**				**θ_h =-3 logit**			
	δ		**P_{ij}**	**p**	**δ**		**P_{ij}**	**p**	**δ**		**P_{ij}**	**p**
1	0	+	0,95		0	+	0,5	0,5	0	–	0,05	0,05
2	0	+	0,95	0,91	0	–	0,5	0,25	0	–	0,05	0,002
3	3	+	0,5	0,5	0	+	0,5	0,5	-3	+	0,5	0,5
4	3	–	0,5	0,25	0	–	0,5	0,25	-3	–	0,5	0,25
5	3	+	0,5	0,5	0	+	0,5	0,5	-3	+	0,5	0,5

	θ_h =3 logit				θ_h =0 logit				θ_h =-3 logit			
	δ		**P_{ij}**	**p**	**δ**		**P_{ij}**	**p**	**δ**		**P_{ij}**	**p**
6	3	–	0,5	0,25	0	–	0,5	0,25	-3	–	0,5	0,25
7	3	+	0,5	0,5	0	+	0,5	0,5	-3	+	0,5	0,5
8	3	–	0,5	0,25	0	–	0,5	0,25	-3	–	0,5	0,25
9	3	+	0,5	0,5	0	+	0,5	0,5	-3	+	0,5	0,5
10	3	–	0,5	0,25	0	–	0,5	0,25	-3	–	0,5	0,25
11	3	+	0,5	0,5	0	+	0,5	0,5	-3	+	0,5	0,5
12	3	–	0,5	0,25	0	–	0,5	0,25	-3	–	0,5	0,25
13	3	+	0,5	0,5	0	+	0,5	0,5	-3	+	0,5	0,5
14	3	–	0,5	0,25	0	–	0,5	0,25	-3	–	0,5	0,25
15	3	+	0,5	0,5	0	+	0,5	0,5	-3	+	0,5	0,5
16	3	–	0,5	0,25	0	–	0,5	0,25	-3	–	0,5	0,25
17	3	+	0,5	0,5	0	+	0,5	0,5	-3	+	0,5	0,5
18	3	–	0,5	0,25	0	–	0,5	0,25	-3	–	0,5	0,25
19	3	+	0,5	0,5	0	+	0,5	0,5	-3	+	0,5	0,5
20	3	–	0,5	0,25	0	–	0,5	0,25	-3	–	0,5	0,25

According to results of modeling it has been received meanings of θ_{je} at different quantity of questions in the test (table 5).

Table 5.

Illustration of difference between the hypothetical (set) level and empirical learning curve at the different quantity of questions in the test, testing with block adaptation

θ_{jh}	n=20	n=15	n=10	n=7
	θ_{je}	**θ_{je}**	**θ_{je}**	**θ_{je}**
3	2,0	2,21	2,21	1,31
0	-0,9	-0,93	-1,01	-1,26
-3	-5,6	-5,2	-4,6	-4,34

For comparison testing with continuous and block adaptation to all hypothetical learning curves, the Euclidean space ε between set of empirical levels for two kinds of testing has been used at the same hypothetical levels:

$$\varepsilon\left(\theta_h, \theta_e\right) = \frac{1}{\sqrt{n}} \cdot \sqrt{\sum_{j=1}^{n}\left(\theta_{je} - \theta_{jh}\right)^2},$$

where n – quantity of gradations learning curve.

In our case, for continuous adaptation $n = 4$, for block adaptation – $n = 3$.

The calculations which were made are presented in the table 6.

Table 6.

Euclidean space between empirical learning curve and hypothetical level for continuous and block adaptation

		n=20	n=15	n=10	n=7
Continuous adaptation	ε_C	0,535	0,564	0,547	0,559
Block adaptation	ε_B	0,047	0,109	0,067	0,238

Calculations of Euclidean space for block adaptation with hypothetical learning curve $\theta_{\Gamma j} = 5$ logit to consider are incorrectly, because the modeling has been done with a small gradation of levels of complexity questions (-3, 0, 3 logit).

According to the data from the table 6 it is visible that block adaptation gives better results, than continuous one.

Findings

Researches algorithms of adaptive testing are executed by means of modeling testing procedure where student's model is presented by hypothetical learning curve, model of the stratified test space – by gradations of levels of complexity, model of the student's reaction on the question – by the Rasch model. Rules of making decision according to algorithms of adaptive testing have been formulated.

The next results of pyramidal testing researches with continuous adaptation were received: at a small quantity of gradation of complexity stratified test space the accuracy of testing results (what is characterized by a difference between empirical and hypothetical levels) is low, but quantity of questions can be small (7-10 questions). That is, the pyramidal testing with continuous adaptation can be used for estimation of a learning curve with low quantity of questions. At "flex level" testing, the quantity of questions should be increased approximately twice.

The test space is divided on difficulty levels. If quantity of these levels is increasing, the accuracy of definition learning curve in testing will increase. As researches have shown, levels of complexity should be divided non-uniformly for increasing accuracy in the middle part of test space complexity range. The increasing quantity levels of complexity gradations in l times demands increase in a minimum quantity of questions in the test in $\sqrt{l}$ times. That is, the minimum quantity of questions in the test will be equal approximately 20.

Modeling of algorithms testing with continuous and block adaptation showed that block adaptation gives better results of testing, than continuous one.

References

1. *Huey-Ing Liu, Min-Num Yang*. QoL Guaranteed Adaptation and Personalization in E-Learning Systems// IEEE Transactions on Education, vol 48, № 4, November 2005 p. 676 – 687.
2. *L. V. Zayceva* Models and methods of adaptation to pupils in systems of computer training // Educational Technology & Society. – Nr. 6(3), 2003. – 456 p.
3. *D. I. Popov* The way of an estimation of knowledge in remote training based on indistinct relations // Remote education, N6, 2000, p. 26 – 29.
4. *U. B. Neyman, V. A. Hlebnikov* Introduction in the theory of modeling and parameterization of pedagogical tests. – Moscow, – 2000 – 168 p.

Manuel Nascimento Dias Barcelos Júnior
Luciano Gonçalves Noleto
Antônio Cesar Pinho Brasil Júnior