The algorithm for knowledge assessment based on the Rusch model

Alexander A. Kostikov[1], Kateryna V. Vlasenko[2,3], Iryna V. Lovianova[4], Sergii V. Volkov[5] and Evgeny O. Avramov[1]

[1]*Donbass State Engineering Academy, 72 Academichna Str., Kramatorsk, 84313, Ukraine*

[2]*National University of "Kyiv Mohyla Academy", 2 Hryhoriya Skovorody Str., Kyiv, 04655, Ukraine*

[3]*Technical University "Metinvest Polytechnic" LLC, 71A Sechenov Str., Mariupol, 87524, Ukraine*

[4]*Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine*

[5]*The Institute of Chemical Technologies of the East Ukrainian Volodymyr Dahl National University, 31 Volodymyrska Str., Rubizhne, 93009, Ukraine*

Abstract

In this paper the algorithm for adaptive testing of students' knowledge in distance learning and an assessment of its effectiveness in the educational process has been proposed. The paper provides an overview of the results of the application of modern test theory, a description and block diagram of the proposed algorithm and the results of its application in the real educational process. The effectiveness of using this algorithm for the objective assessment of students' knowledge has been experimentally shown.

Keywords

adaptive algorithm, Rasch model, Item Response Theory (IRT), information function of test item, latent variables

1. Introduction

1.1. Problem statement

Modern approaches to assessing students' academic achievements are based on the use of classical testing theory and Item Response Theory (IRT). The mathematical background of pedagogical measurement theory was created by Andersen [1, 2], Andrich [3], Avanesov [4], Birnbaum [5], Guttman [6], Linacre [7], Lord et al. [8], Maslak et al. [9], Masters [10], Rasch [11] and other scientists. In IRT, the concept of a latent variable is used. The term "latent variable (parameter)" is usually understood as a theoretical concept that characterizes a certain

hidden property or quality (for example, the level of students' ability, the difficulty of the test task), which cannot be directly measured. The advantages of the classical testing theory are the provision of information about the indicators of the knowledge quality of the subjects, the clarity of the performed calculations and the simple interpretation of the processing data. The main disadvantage is the dependence of the results of evaluating the participants' parameters on the difficulty of the proposed tasks. Application of IRT, based on Rush's models, provides the possibility of the evaluation independence of the latent parameter "ability level" calculated values of participants θ_i from the values of the "item difficulty" β_i. This helps to increase the objectivity of the obtained assessments of the students' ability level and allows to build effective algorithms for assessing knowledge.

The *purpose* of this paper is to develop an algorithm of adaptive testing for objective assessment of students' knowledge in distance learning, which becomes especially relevant in the quarantine of COVID-19.

1.2. State of arts and review

The educational standards of the new generation are based on a competency-based approach to assessing the quality of a student's training, when it is not his knowledge that is tested, first of all, but his readiness to apply it in practice and to act productively in a non-standard situation, the ability to create the required mode of action. Therefore, the quality of training is understood as the degree of the student's readiness to demonstrate the relevant competencies.

The generalization of the world experience in the implementation of the competence-based approach to assessing learning outcomes allows us to make the following conclusions that determine the main approaches to assessing the level of competence mastery, the main of which are the following:

- competencies are dynamic, since they are not an invariable quality in the structure of a pupil's personality, but are able to develop, improve or completely disappear in the absence of an incentive to manifest them. Therefore, we can talk about the level of competence, assess it quantitatively, and monitor it.
- when assessing learning outcomes, it is necessary to consider them in dynamics, which requires diagnostics of the educational process using monitoring procedures.
- the level of possession of a competence is a hidden (latent) parameter of the pupil and direct measurement is not amenable. It can be estimated with a certain probability. Therefore, when evaluating it, a probabilistic approach should be used.

It follows from this that in order to create tools for the automated assessment of the learning outcomes, it is necessary, first of all, to solve two problems:

1) develop theoretical and methodological foundations for modeling and parameterization of the learning process and the diagnostic tools used to evaluate its results;
2) theoretically substantiate and implement software-algorithmic means for processing the results of participants' diagnostics (testing, questionnaires), as well as tools for assessing learning outcomes and the quality of diagnostic tools.

The theoretical and methodological basis for solving these problems was the study results, first of all, by Brown [12], Cronbach [13], Guilford [14], Gulliksen [15], Guttman [6], Kuder and Richardson [16], Luce and Tukey [17], Lord et al. [8], Sax [18], Spearman [19]. They developed the theoretical foundations for the creation of diagnostic materials and the classical approach to processing, analysis and interpretation of diagnostic results: the conceptual apparatus of the classical test theory , criteria and indicators of the quality of diagnostic tools, methodological basics of their design and quality expertise. The issues of scaling and comparison of processing data have been deeply investigated.

The theoretical basis for the creation of tools for automatic assessment of the results of the educational process has received its further development due to the creation of the IRT (Item Response Theory), the foundations of which are set out by Andrich [3, 20], Bezruczko [21], Bond and Fox [22], Bond et al. [23], Eckes [24], Fischer and Molenaar [25], Andrich et al. [26], Ingebo [27], Kim and Baker [28], Lazarsfeld [29], van der Linden and Hambleton [30], Lord [31], Luce and Tukey [17], Perline et al. [32], Smith and Smith [33], Rasch [11], Wilson [34], Wright [35], Wright and Masters [36], Wright and Stone [37], Wright and Linacre [38].

2. Algorithm of adaptive testing based on Rasch model

Adaptive testing is a type of testing in which the order of presentation of test items and the difficulty of the next task depends on the participant's answers to previous items. The basis of adaptive testing systems are statistical models. Very easy and very difficult tasks are automatically uninformative. Therefore, for most tests, the optimal level of difficulty is the item, to which the correct answer is given by about half of the test participants.

The difficulties of the test items is determined experimentally, and the measurement process consists of determining the percentage of participants who are able to give the correct answer to the task in previous experiments.

The problem of developing adaptive algorithms has been considered by Al-A'ali [39], Weiss [40, 41].

The Rush model was used to construct the adaptive testing algorithm. This model is defined by formulas:

$$P_{ni} = \frac{\exp(\theta_n - \beta_i)}{1 + \exp(\theta_n - \beta_i)} \tag{1}$$

where P_{ni} is the probability that the participant $n, n = 1, ..., N$ with the ability θ_n correctly performs the task $i, i = 1, ..., I$, with the difficulty β_i. To start the algorithm it is necessary to determine the initial levels of difficulties. To this end, at the beginning of the testing session the accumulation of primary information about the level of preparation of the participant is carried out. To do this, participant receive N_p tasks with an average level of difficulty. Tasks to determine the initial level of the participant are chosen by the teacher. Then, using the received answers, the initial estimation of the ability level of the student is calculated, and also recalculation of the difficulty level current values of test items is carried out.

The initial assessment of the ability level of the i-th student (in logs) is based on the formula:

$$\theta_i^0 = \ln\left(\frac{p_i}{q_i}\right), \quad i = 1, 2 \cdots N, \tag{2}$$

where N is the number of test participants, p_i is the proportion of correct answers of the i-th participant to all tasks, q_i is the proportion of incorrect answers ($q_i = 1 - p_i$).

The difficulty level of test items in logs is determined by the formula:

$$\beta_j^0 = \ln\left(\frac{q_j}{p_j}\right), \quad j = 1, 2 \cdots M, \tag{3}$$

where M is the number of test items, p_j is the proportion of correct answers of all participants to the j-th test item, q_j is the proportion of incorrect answers.

At the next stage, the initial values in the logs of the ability level of participants θ_i^0 and the initial values in the logs of the difficulty level of the test item β_j^0 are reduced to a same interval scale [8]. The formula for such transition is based on the idea of reducing the impact of the items difficulty on the assessments of test participants.

Pre-calculating the average value of the initial logits of the students' ability level

$$\bar{\theta} = \frac{\sum\limits_{i=1}^{N} \theta_i^0}{N}$$

and the standard deviation V of the initial values distribution of the parameter θ

$$V^2 = \frac{\sum\limits_{i=1}^{N} \left(\theta_i^0 - \bar{\theta}\right)^2}{N - 1},$$

we obtain a formula for calculating the dofficulty level logit of the j-th item

$$\beta_j = \bar{\theta} + Y \cdot \beta_j^0, \quad j = \overline{1, M}, \tag{4}$$

where

$$Y = \sqrt{1 + \frac{V^2}{2.89}}$$

Similarly, calculating

$$\bar{\beta} = \frac{\sum\limits_{j=1}^{M} \beta_j^0}{M}, \quad W = \sqrt{\frac{\sum\limits_{j}^{M} \left(\beta_j^0 - \bar{\beta}\right)^2}{M - 1}}$$

we get the formula for calculating the ability level logit of the i-th student:

$$\theta_i = \bar{\beta} + X \cdot \theta_i^0, \quad i = \overline{1, N}, \tag{5}$$

where $X = \left(1 + \frac{W^2}{2.89}\right)^{\frac{1}{2}}$.

The obtained values allow to compare the level of students' ability with the level of test item difficulty. If $\theta_i - \beta_j$ is a negative quantity and is large in modulus, then the problem of difficulty β_j is too difficult for a student with the ability level θ_i, and it will not be useful for measuring the level of knowledge of the i-th student. If this difference is positive and large in modulus, then

the task is too easy, it has long been mastered by the student. If $\theta_i = \beta_j$, then the probability that the student correctly completes the task is equal to 0.5.

The information function of the i-th problem for the Rush model (1) $I_i(\theta)$ is defined as the product of the probability of the correct answer $P_i(\theta)$ to this problem on the probability of the incorrect answer $Q_i(\theta)$ [8]

$$I_i(\theta) = P_i(\theta) \cdot Q_i(\theta) \tag{6}$$

Figure 1 shows the information function of the i-th item.
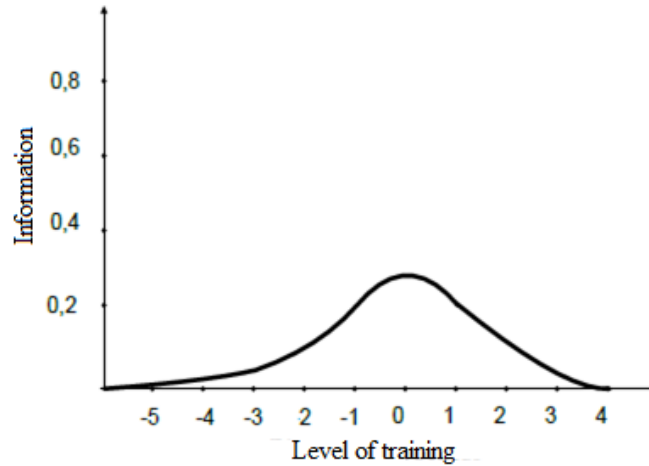
Figure 1: Information function of the test task.

Figure 1 shows that the test item, the answer to which all students know, does not provide any information, as well as the item, the answer to which no one knows. We get useful information when some participants know the answer to the task and some do not.

The information function of the test is calculated as the sum of the information functions of the test items [8]:

$$I(\theta) = D^2 \cdot \sum_{j=1}^{M} I_j(\theta) \tag{7}$$

where D is the correction factor ($D = 1.7$), necessary to approximate the distribution of logistic probability to the law of normal distribution.

After calculating the information function, the measurement error SE is calculated, the value of which is used to check the condition of the end of the test procedure. In the Rusch model, the measurement error depends on the level of training θ and is calculated by the formula [8]:

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \tag{8}$$

If the error takes a value less than the threshold set by the teacher, the adaptive testing algorithm ends. Otherwise, the following test task is selected. To select the next task, use the value of θ_i, calculated by formula (5). The next task is the one whose difficulty level is

closest to the current assessment of the ability level of the participant. This task has the largest information contribution and its choice reduces the total number of required test tasks.

Thus, the developed adaptive testing algorithm consists of the following stages:

1. Selection of 5 tasks of average difficulty from the bank of questions, which is determined by the teacher.
2. Finding the initial level of student's ability θ_i^0 and the initial difficulty level of items β_j^0 by formulas (2) and (3).
3. Reduction of the obtained initial values θ_i^0 and β_j^0 to a single interval scale using formulas (5) and (4).
4. Calculation of the information function of test tasks to which the student answered by formulas (6) and (7).
5. Finding the measurement error by the formula (8).
6. If the measurement error is less than the threshold, the adaptive testing is completed.
7. If not, then the next task is selected from the condition $\left| \theta_i - \beta_j \right| = \min$.
8. Then the algorithm is repeated starting from point 3.

The block diagram of the algorithm is shown in figure 2.

3. Results

Let us consider the procedure for calculating the parameters of student ability level θ_i and item difficulty parameter β_i from empirical data. As initial data we will take results of testing of students in Moodle system on discipline "Higher Mathematics" of the Mathematics and Modeling Departement of the Donbass State Engineering Academy (Table 1). Table 1 shows the records of the first 10 test participants. A total of 50 participants took part in the testing.

Table 1
Test results in the Moodle system in the discipline "Higher Mathematics" of the Mathematics and Modeling Department of the Donbass State Engineering Academy

Participant's number	Score	Number of correct answers	p_i	q_i	θ_i^0
1	90	18	0.9	0.1	2.197225
2	75	15	0.75	0.25	1.098612
3	85	17	0.85	0.15	1.734601
4	100	20	1	0	∞
5	75	15	0.75	0.25	1.098612
6	100	20	1	0	∞
7	90	18	0.9	0.1	2.197225
8	90	18	0.9	0.1	2.197225
9	70	14	0.7	0.3	0.847298
10	85	17	0.85	0.15	1.734601

The test in this discipline consisted of 20 questions. First, it is necessary to calculate the proportions of correct p_i and incorrect q_i answers of participants. These values are calculated
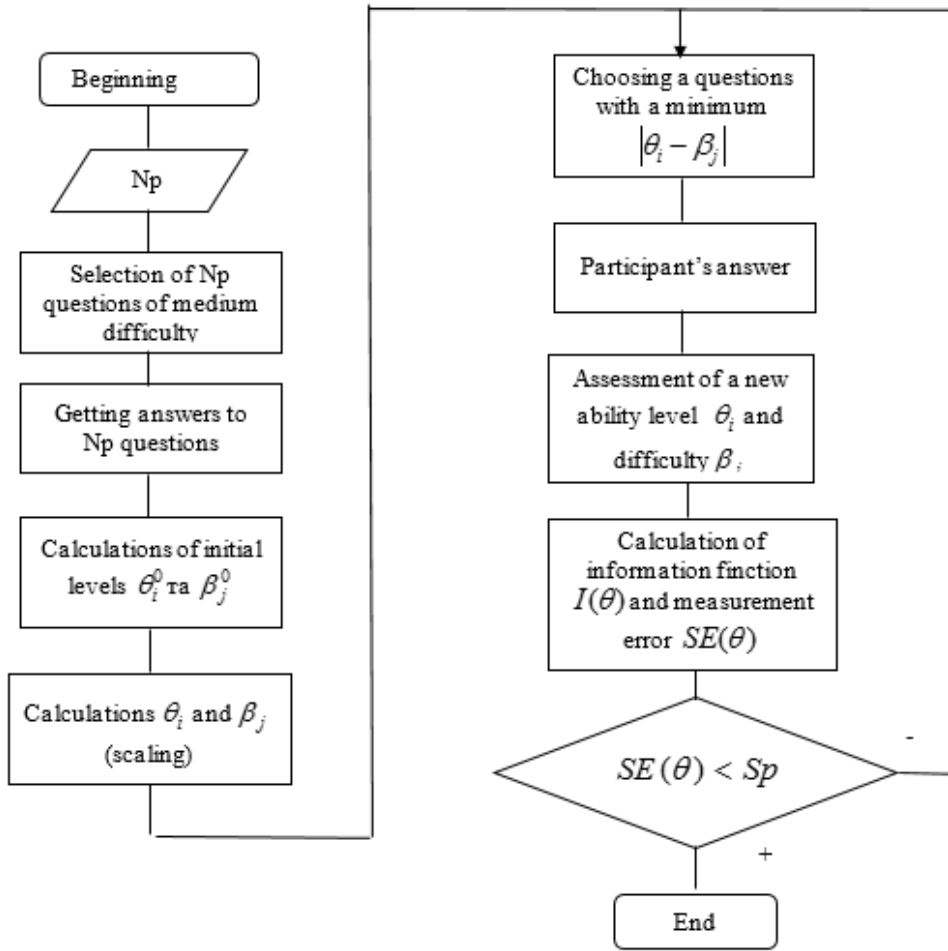
Figure 2: Block diagram of the adaptive testing algorithm.

by formulas

$$p_i = \frac{R_i}{N}, q_i = 1 - p_i, \tag{9}$$

where R_i is the number of correct answers for the i-th test item, $= 1, 2, ..., n$, and n is the number of items in the test.

For example, for the first participant of testing we have

$$p_1 = \frac{18}{20} = 0.9 \ q_1 = 1 - 0.9 = 0.1$$

The values p_i and q_i are given in columns 3 and 4 of table 1.

Next, calculate the initial values θ_1^0 of the ability level of participants by formula (2). For the first participant we have

$$\theta_1^0 = \ln \frac{0.9}{0.1} = 2.197$$

Using the statistical module Moodle, the following characteristics were obtained for test tasks: facility index(F), standard deviation (SD), random guess score (RGS), intended weight, effective weight, distinction, distinction efficiency. These data are shown in table 2.

Table 2
Statistical characteristics obtained using the statistical module of the Moodle system based on the results of final testing in the discipline "Higher Mathematics"

Q#	F	SD	RGS	Intended weight	Effective weight	Distinction	Distinguishing efficiency
1	98.00%	14.14%	33.33%	5.00%		-11.54%	-28.62%
2	94.00%	23.99%	33.33%	5.00%	3.41%	6.93%	11.28%
3	90.00%	30.30%	16.67%	5.00%	6.75%	44.07%	65.85%
4	94.00%	23.99%	20.00%	5.00%	4.66%	22.91%	39.11%
5	96.00%	19.79%	20.00%	5.00%	3.34%	11.72%	23.66%
6	90.00%	30.30%	14.29%	5.00%	3.18%	-1.53%	-2.22%
7	92.00%	27.40%	14.29%	5.00%	6.32%	43.38%	70.79%
8	84.00%	37.03%	20.00%	5.00%	6.48%	26.08%	35.44%
9	88.00%	32.83%	20.00%	5.00%	5.32%	17.26%	23.76%
10	74.00%	44.31%	20.00%	5.00%	9.75%	68.31%	84.84%
11	98.00%	14.14%	20.00%	5.00%	2.85%	14.64%	35.69%
12	100.00%	0.00%	16.67%	5.00%	0.00%		
13	94.00%	23.99%	33.33%	5.00%	4.93%	27.00%	45.87%
14	90.00%	30.30%	33.33%	5.00%	5.51%	23.81%	34.88%
15	88.00%	32.83%	25.00%	5.00%	5.32%	17.26%	23.76%
16	90.00%	30.30%	33.33%	5.00%	5.51%	23.81%	33.33%
17	42.00%	49.86%	20.00%	5.00%	5.23%	-2.60%	-3.57%
18	80.00%	40.41%	33.33%	5.00%	8.11%	45.46%	56.25%
19	56.00%	50.14%	20.00%	5.00%	7.01%	13.80%	17.23%
20	82.00%	38.81%	20.00%	5.00%	6.32%	21.10%	27.68%

Based on the data in table 2, we can estimate the initial values of the item difficulty parameter. By formula (3) for the first problem we obtain

$$\beta_1^0 = \ln \frac{2}{98} = -3.891$$

The results of calculations of the initial values of the item difficulty parameter are given in table 3.

As can be seen from table 3, all participants in the quiz answered the 12th item, so the score was equal to infinity with a minus sign. But practically at $\beta > -6$ the probability value $P_i(\beta)$ close to one. These items are performed by all participants and they become redundant. Items with $\beta > 6$ are also useless. Such items will not be overcome by any participant and they do not carry any information about differences in the students' ability levels.

In tables 1 and 3, the parameter values θ_i^0 and β_i^0 are on different interval scales. In order to reduce them to a single scale of standard estimates, it is necessary to calculate the variances V^2 and W^2 using the data from tables 1 and 3. Infinite data are excluded from consideration.

Table 3
Initial values β_i^0 of the item difficulty parameter

Q#	Progress	p_i	q_i	β_i^0
1	98.00%	0.98	0.02	-3.89182
2	94.00%	0.94	0.06	-2.75154
3	90.00%	0.90	0.10	-2.19722
4	94.00%	0.94	0.06	-2.75154
5	96.00%	0.96	0.04	-3.17805
6	90.00%	0.90	0.10	-2.19722
7	92.00%	0.92	0.08	-2.44235
8	84.00%	0.84	0.16	-1.65823
9	88.00%	0.88	0.12	-1.99243
10	74,00%	0.74	0.26	-1.04597
11	98.00%	0.98	0.02	-3.89182
12	100.00%	1.00	0.00	$-\infty$
13	94.00%	0.94	0.06	-2.75154
14	90.00%	0.90	0.10	-2.19722
15	88.00%	0.88	0.12	-1.99243
16	90.00%	0.90	0.10	-2.19722
17	42,00%	0.42	0.58	0.322773
18	80.00%	0.80	0.20	-1.38629
19	56,00%	0.56	0.44	-0.24116
20	82.00%	0.82	0.18	-1.51635

Calculating the variance, we obtain

$$V^2 = \frac{\sum_{i=1}^{N} \left(\theta_i^0 - \bar{\theta}\right)^2}{N - 1} = 0.634,$$

$$W^2 = \frac{\sum_{j}^{M} \left(\beta_j^0 - \bar{\beta}\right)^2}{M - 1} = 4.873$$

Next, we calculate the angular coefficients [8]:

$$Y = \sqrt{1 + \frac{V^2}{2.89}} = 1.104$$

$$X = \sqrt{1 + \frac{W^2}{2.89}} = 1.63$$

Next on the formulas

$$\theta_i = -2.103 + 1.104\theta_i^0$$

$$\beta_i = 1.86 + 1.63\beta_i^0$$

calculate the scaled values β_i and θ_i.

In tables 4 and 5 scaled parameter values are provided.

Table 4
Scaled values of item difficulty parameter β_i

Q#	β_i^0	β_i
1	-3.89182	-4.48367
2	-2.75154	-2.625
3	-2.19722	-1.72148
4	-2.75154	-2,625
5	-3.17805	-3.32023
6	-2.19722	-1.72148
7	-2.44235	-2.12103
8	-1.65823	-0.84291
9	-1.99243	-1.38766
10	-1.04597	0.155071
11	-3.89182	-4.48367
13	-2.75154	-2,625
14	-2.19722	-1.72148
15	-1.99243	-1.38766
16	-2.19722	-1.72148
17	0.322773	2.386121
18	-1.38629	-0.39966
19	-0.24116	1.466906
20	-1.51635	-0.61165

Table 5
Scaled values of the ability level θ_i

Participant's number	θ_i^0	θ_i
1	2.197225	0.322736
2	1.098612	-0.89013
3	1.734601	-0.188
5	1.098612	-0.89013
7	2.197225	0.322736
8	2.197225	0.322736
9	0.847298	-1.16758
10	1,734601	-0.188

The sum of the scaled difficulty levels of test items is -27.93.

This means that the test items are very easy. This test is not balanced, it contains a lot of easy items. It is necessary to strive to ensure that this amount is close to zero. Thus, the assessment of latent parameters allows to determine noninformative items that should be excluded from the quiz. The use of the developed adaptive algorithm will allow to objectively assess the level of students' knowledge.

The graph of the information function of test items and the test as a whole, defined by formulas (6) and (7), is shown in figure 3.

Figure 3 shows that the information function has one clearly expressed maximum. This is a
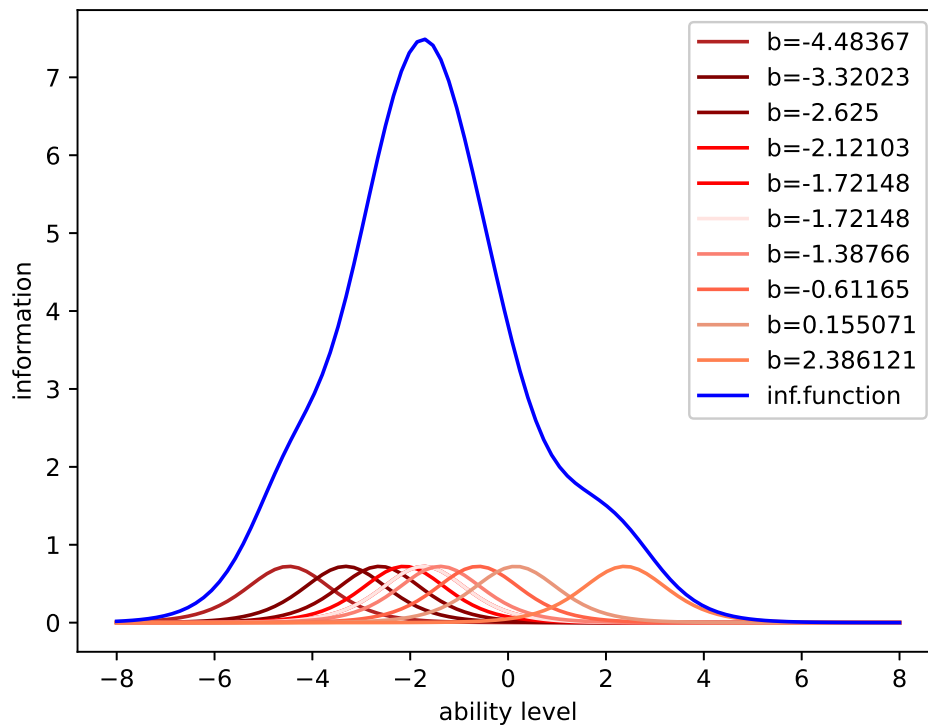
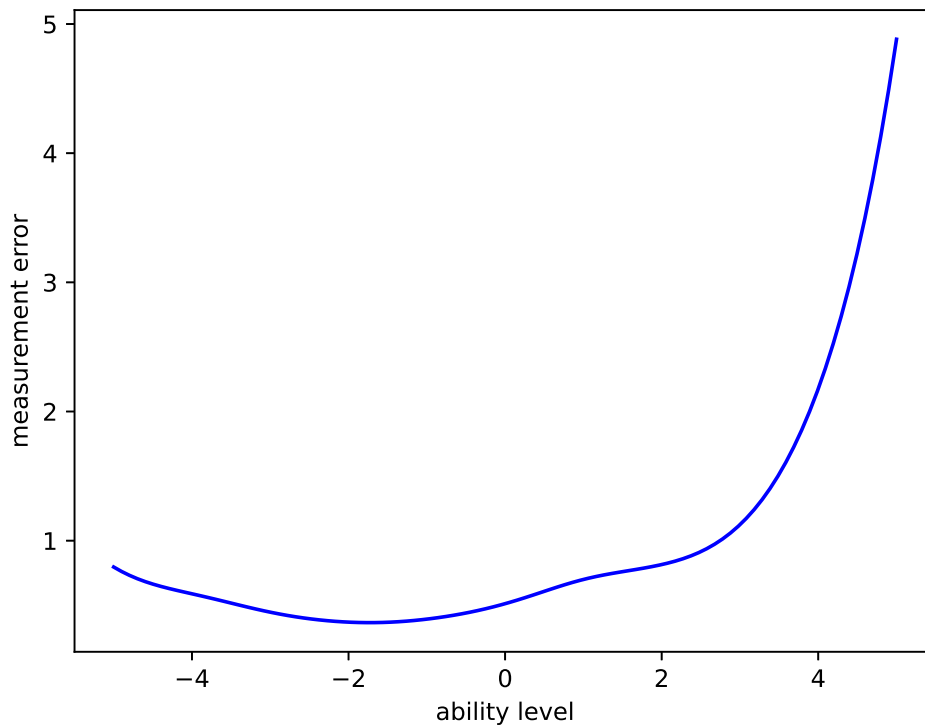**Figure 3:** Information functions of the test and test items.

sign of a "good" test. However, it can be seen that the test contains a lot of easy test items with difficulties in the interval (-3; -2), which can be excluded from the test. Also in the test there are many easy tasks with the same difficulties, which can also be excluded from the test without violation of its information content. However, the more difficult tasks (with difficulties of 1-2 logits) are clearly not enough in the test, so it is necessary to add more complex tasks.

The graph of the measurement error, depending on the level of training, is shown in figure 4.

It can be seen from the graph that the measurement error is large for the values of the ability in the interval (2,4), which is associated with the lack of test items of increased difficulty.

4. Discussion

The purpose of this paper was to automate the process of testing students' knowledge, which is especially relevant for distance learning. To achieve this goal, an adaptive testing algorithm based on the Rush model was proposed and the modeling of the students' knowledge assessment process using this algorithm was carried out. The results of testing their knowledge in the course "Higher Mathematics" obtained in the Moodle system were taken as the initial values of the tasks complexity and the levels of the students' ability.

Figure 4: Measurement error graph.

As a result of modeling, the levels of students' abilities were recalculated, the information functions of the test tasks and the entire test as a whole were built, the standard measurement error was calculated, depending on the student's ability level. The analysis of the obtained results allows us to conclude that the test is not balanced, contains too many easy tasks. In this case, these are tasks with numbers 1, 3, 11. Removing them from the test will reduce the number of test items and speed up the process of determining the student's level of training. A change in the assessment of the student's ability level as a result of testing indicates the need to introduce an adaptive testing system into the educational process, which will improve the quality of assessment of student knowledge.

These conclusions are confirmed by the works of other authors. So, Al-A'ali [39] shown that the use of adaptive testing based on IRT made it possible to reduce the number of test tasks and increase the reliability of determining the level of student readiness. The effectiveness of the use of adaptive testing to improve the quality of pedagogical measurements is evidenced by Weiss [40, 41].

5. Conclusions

As a result of this work, the following results were obtained:

1. An algorithm of adaptive knowledge assessment based on the IRT approaches was proposed. This algorithm consists of an initial assessment of the difficulty level of test items and students' abilities, scaling of these parameters, selection of the next question based on minimizing the module of their difference and estimation of the measuring error of the knowledge level by the information function of the proposed question.
2. The test parameters were evaluated on the basis of IRT theory, which identified non-informative test questions that should be excluded from the set of test items.

The results of the study showed the effectiveness of using IRT to assess knowledge.

References

[1] E. B. Andersen, The asymptotic distribution of conditional likelihood ratio tests, Journal of the American Statistical Association 66 (1971) 630–633. doi:10.1080/01621459.1971.10482321.

[2] E. B. Andersen, A goodness of fit test for the Rasch model, Psychometrika 38 (1973) 123–140. doi:10.1007/BF02291180.

[3] D. Andrich, Rasch Models for Measurement, Thousand Oaks, 2021. URL: https://methods.sagepub.com/book/rasch-models-for-measurement. doi:10.4135/9781412985598.

[4] V. S. Avanesov, The problem of psychological tests, Soviet Education 22 (1980) 6–23. doi:10.2753/RES1060-939322066.

[5] A. Birnbaum, Combining independent tests of significance*, Journal of the American Statistical Association 49 (1954) 559–574. doi:10.1080/01621459.1954.10483521.

[6] L. Guttman, A basis for scaling qualitative data, American Sociological Review 9 (1944) 139–150. doi:10.2307/2086306.

[7] J. M. Linacre, Predicting responses from Rasch measures, Journal of Applied Measurement 11 (2010) 1–10.

[8] F. M. Lord, M. R. Novick, A. Birnbaum, Statistical theories of mental test scores, Addison-Wesley, Oxford, 1968.

[9] A. A. Maslak, G. Karabatsos, T. S. Anisimova, S. A. Osipov, Measuring and comparing higher education quality between countries worldwide, Journal of Applied Measurement 6 (2005) 432–442.

[10] G. N. Masters, Educational measurement: Prospects for research and innovation, The Australian Educational Researcher 15 (1988) 23–34. doi:10.1007/BF03219425.

[11] G. Rasch, Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests, Nielsen & Lydiche, 1960.

[12] W. Brown, Some experimental results in the correlation of mental abilities1, British Journal of Psychology, 1904-1920 3 (1910) 296–322. doi:10.1111/j.2044-8295.1910.tb00207.x.

[13] L. J. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika 16 (1951) 297–334. doi:10.1007/BF02310555.

[14] J. P. Guilford, Fundamental statistics in psychology and education, McGraw-Hill, New York, 1942.

[15] H. Gulliksen, Perspective on Educational Measurement, Applied Psychological Measurement 10 (1986) 109–132. doi:10.1177/014662168601000201.

[16] G. F. Kuder, M. W. Richardson, The theory of the estimation of test reliability, Psychometrika 2 (1937) 151–160. doi:10.1007/BF02288391.

[17] R. D. Luce, J. W. Tukey, Simultaneous conjoint measurement: A new type of fundamental measurement, Journal of Mathematical Psychology 1 (1964) 1–27. doi:10.1016/0022-2496(64)90015-X.

[18] G. Sax, Principles of educational and psychological measurement and evaluation, 3rd ed., Wadsworth Pub. Co., Belmont, 1989.

[19] C. Spearman, Correlation calculated from faulty data, British Journal of Psychology, 1904-1920 3 (1910) 271–295. doi:10.1111/j.2044-8295.1910.tb00206.x.

[20] D. Andrich, The Rasch model explained, in: Applied Rasch measurement: A book of exemplars, Springer, 2005, pp. 27–59.

[21] N. Bezruczko (Ed.), Rasch measurement in health sciences, Jam Press Maple Grove, MN, 2005.

[22] T. Bond, C. Fox, Applying the Rasch model: Fundamental measurement in the human sciences, second ed., 2007. doi:10.4324/9781410614575.

[23] T. Bond, Z. Yan, M. Heene, Applying the Rasch model: Fundamental measurement in the human sciences, fourth ed., Routledge, 2020.

[24] T. Eckes, Introduction to Many-Facet Rasch Measurement, Peter Lang, Bern, Switzerland, 2011. URL: https://www.peterlang.com/view/title/13347.

[25] G. H. Fischer, I. W. Molenaar (Eds.), Rasch models: Foundations, recent developments, and applications, Springer Science & Business Media, 1995. doi:10.1007/978-1-4612-4230-7.

[26] D. Andrich, B. Sheridan, G. Luo, Rumm2010: Rasch unidimensional measurement models, 2001. URL: http://www.rummlab.com.au/.

[27] G. S. Ingebo, Probability in the Measure of Achievement, Mesa Press, 1997.

[28] S.-H. Kim, F. B. Baker, birtr: A Package for "The Basics of Item Response Theory Using R", Applied Psychological Measurement 42 (2018) 403–404. doi:10.1177/0146621617748327.

[29] P. F. Lazarsfeld, Regression analysis with dichotomous attributes, Social Science Research 1 (1972) 25–34. doi:10.1016/0049-089X(72)90056-7.

[30] W. J. van der Linden, R. K. Hambleton (Eds.), Handbook of Modern Item Response Theory, Springer Science & Business Media, 1997. doi:10.1007/978-1-4757-2691-6.

[31] F. M. Lord, Applications of Item Response Theory To Practical Testing Problems, Routledge, 1980. doi:10.4324/9780203056615.

[32] R. Perline, B. D. Wright, H. Wainer, The Rasch model as additive conjoint measurement, Applied Psychological Measurement 3 (1979) 237–255. doi:10.1177/014662167900300213.

[33] E. V. Smith, R. M. Smith (Eds.), Introduction to Rasch measurement: Theory, models and applications, JAM Press, 2004.

[34] M. Wilson, Constructing Measures: An Item Response Modeling Approach, Routledge, 2005.

[35] B. D. Wright, Solving Measurement Problems with the Rasch Model, Journal of Educational Measurement 14 (1977) 97–116. URL: http://www.jstor.org/stable/1434010.

[36] B. D. Wright, G. N. Masters, Rating scale analysis, Mesa Press, Chicago, 1982.

[37] B. D. Wright, M. H. Stone, Best test design, Mesa Press, Chicago, 1979. URL: https://www.rasch.org/BTD_RSA/pdf%20[reduced%20size]/Best%20Test%20Design.pdf.

[38] B. D. Wright, J. M. Linacre, Dichotomous rasch model derived from specific objectivity, Rasch measurement transactions 1 (1987) 5–6. URL: https://www.rasch.org/rmt/rmt11a.htm.

[39] M. Al-A'ali, IRT-Item Response Theory Assessment for an Adaptive Teaching Assessment System, in: Proceedings of the 10th WSEAS International Conference on APPLIED MATH-EMATICS, MATH'06, World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, 2006, p. 518–522.

[40] D. J. Weiss, Improving measurement quality and efficiency with adaptive testing, Applied Psychological Measurement 6 (1982) 473–492. doi:10.1177/014662168200600408.

[41] D. J. Weiss, Computerized adaptive testing for effective and efficient measurement in counseling and education, Measurement and Evaluation in Counseling and Development 37 (2004) 70–84. doi:10.1080/07481756.2004.11909751.