

Pedagogical Diagnostics with Use of Computer Technologies

Lyudmyla Bilousova¹, Oleksandr Kolgatin¹ and Larisa Kolgatina¹

¹Kharkiv National Pedagogical University named after G.S.Skovoroda, Kharkiv, Ukraine

lib215@list.ru, kolgatin@ukr.net, larakl@ukr.net

Abstract. The technology of the automated pedagogical diagnostics is analysed. The testing strategy, oriented for pedagogical diagnostics purpose, and grading algorithm, which corresponds to Ukrainian school grading standards, are suggested. "Expert 3.05" software for automated pedagogical testing is designed. The methods of administration of the database of the test items are proposed. Some tests on the mathematical topics are prepared with "Expert 3.05". The approbation of these tests in the educational process of Kharkov National Pedagogical University named after G.S.Skovoroda is analysed.

Keywords. E-learning, Diagnostics, Test

Key terms. InformationCommunicationTechnology, Teaching Process

1 Introduction

Pedagogical diagnostics is the integral part of adaptive E-learning courses. The unconditional quality of testing is its high informative abilities. However, in practice the large part of the test information is often not used. Computer technologies give us possibility to organize the qualitative pedagogical diagnostics at new level. Modern automated systems which can be qualified as expert systems are capable to supply comprehensive algorithms of testing and analysis of the test results. Testing with use of computers allows a teacher to obtain the summary characteristics of knowledge and skills of the pupils' group and to use this information to choose the teaching methods. A study of such algorithms is a wide field of the scientific work. Therefore, the aim of our paper is to design methods of the pedagogical diagnostics, which satisfy following demands:

- Different forms of the intellectual activities of an examinee are attracted in process of testing;
- The automated system of the pedagogical diagnostics ensures its diagnostic abilities at wide differences of the examinees mastering;
- Processing of the test results provides maximum information for an examinee and a teacher to correct the educational process.

2 Objectives

The first stage of pedagogical diagnostics organising is a construction of an idealised pedagogical model that is allocation of basic elements of knowledge and skills, as well as detecting the level of its mastering.

The second stage represents creation of the problems system which covers all elements of knowledge and skills and all levels of their mastering.

We cannot design test as a system of test items of equal difficulty, in spite of recommendation of the classic test theory. Such approach gave the best tests for discrimination of examinees into several groups. However, the test with equal items has low validity for examinees with bad mastering because of guessing answers. Validity of such a test is also low for good mastering examinees because of lack of attention. Therefore, it is certainly necessary to include problems of different difficulty to the test.

How to design a test item of advanced difficulty? What is difficulty? Why the most of examinees do not solve some problems?

We cannot use problems of the reproductive level as items of advanced difficulty. There are not difficult facts and easy facts. Our educational process should be organised to provide steady knowledge of all compulsory facts. If the most of examinees do not know some compulsory facts, it means that we should correct our teaching. We are against using items which correspond to facts that are fragmentary studied and which are not basic for the tested topic. Therefore, all problems corresponding to the reproductive knowledge must have equal difficulty.

Someone can increase the difficulty of an item by combining several operations in this item. Such approach leads to increasing influence of lack of attention on the test results, as well as to necessity of using weight coefficients and to decreasing of the measuring accuracy of the test. We are also against using problems which correspond to facts that are fragmentary studied and do not form basis of the topic being tested.

In our opinion, the item of advanced difficulty should be connected with use of more difficult, not reproductive kinds of the intellectual activities [1], [2].

Full and high-qualitative pedagogical diagnostics should be built on the system of test items of all levels: reproductive and productive. By analogy with levels of educational achievements [2], which are standardised by the Ukrainian Ministry of Education and Science [3] we propose the following levels of the test items:

1. Initial level - it is the very simple problems which assume the reproductive character of the student's activities, mainly distinguishing. The difficulty index of these test items is about 1, the most of the examined students execute these items correctly.
2. Average level - it is the problems which assume the reproductive activities, these problems cover all basic facts and unary skills according to curriculum. A database of items of this level is designed the most naturally. According to the Ukrainian standards [3], the student can continue education, if he (she) knows not less than 50% of compulsory facts determined by curriculum. Therefore, by linear estimation, average index of difficulty must be near 75% for the reproductive items.

3. Sufficient level - these items assume the examinee applies his knowledge and skills for solving problems in standard situation.
4. High level - these test items are practical problems which assume executing of new algorithm, carrying knowledge into new, non-standard situation, etc. These items can lose creative nature, if the method of its solving was explained in the process of learning. Therefore, the database of the items of the level 4 requires continuous analysis and modernisation.

We propose the vector processing of the test results – separate calculation of the score for items of every level. It allows to avoid the use of the artificial weight coefficients and to provide the comprehensive algorithm of adaptive strategy of testing and grading. We also propose the separate processing of the results for the test items according the elements of knowledge and skills.

Using computer for test administration allows to analyse the examinee's results directly in process of testing and to suggest an examinee the items, which mostly correspond to his (her) level of educational achievements. Such approach is often called adaptive or quasi-adaptive testing [4].

3 Model and Algorithm

A choice of the items level for start of testing is an important question of the adaptive strategy. Testing usually starts from the simplest items. Such approach provides decreasing of psychological discomfort and creates the atmosphere of competition, the feeling of growth according to complication of the problems. Taking into account this consideration we propose to start testing just from items of the level 2 which are the simplest for the examinees that will obtain positive grade.

There is additional argument to choose the level 2 as the start level of testing. The test items of the level 2 reflect the compulsory facts of the study topic; these problems cannot be excluded from the test process. It is not worthwhile to start test from the items of the level 3, because productive and, especially, creative problems are based on sufficiently wide spectrum of knowledge, and it is not always possible to detect, which exactly element of curriculum is not mastered by an examinee. The items of the level 1 are intended for students whose mastering is not satisfactory; therefore, there is no need to suggest these items to all examinees. Our testing strategy and algorithm of grading are presented on the fig. 1.

Here are some comments to fig. 1. The testing starts with the items of the level 2. An examinee solves the compulsory minimum of the items on the level 2, automated system calculates S_2 - his (her) score on the level 2 and estimates the error of the score. If accuracy is enough, the automated system chooses a grade or increases the level of items, which are being suggested to the examinee. Otherwise, the items of the level 2 are being suggested to the examinee until the accuracy become satisfactory. It should be underlined that accuracy depends not only on the number of items, but it depends on the individual test score [5]. The necessary accuracy is also conditioned

by the differences between the test score and the key points for decision about grading or a rise of a level.

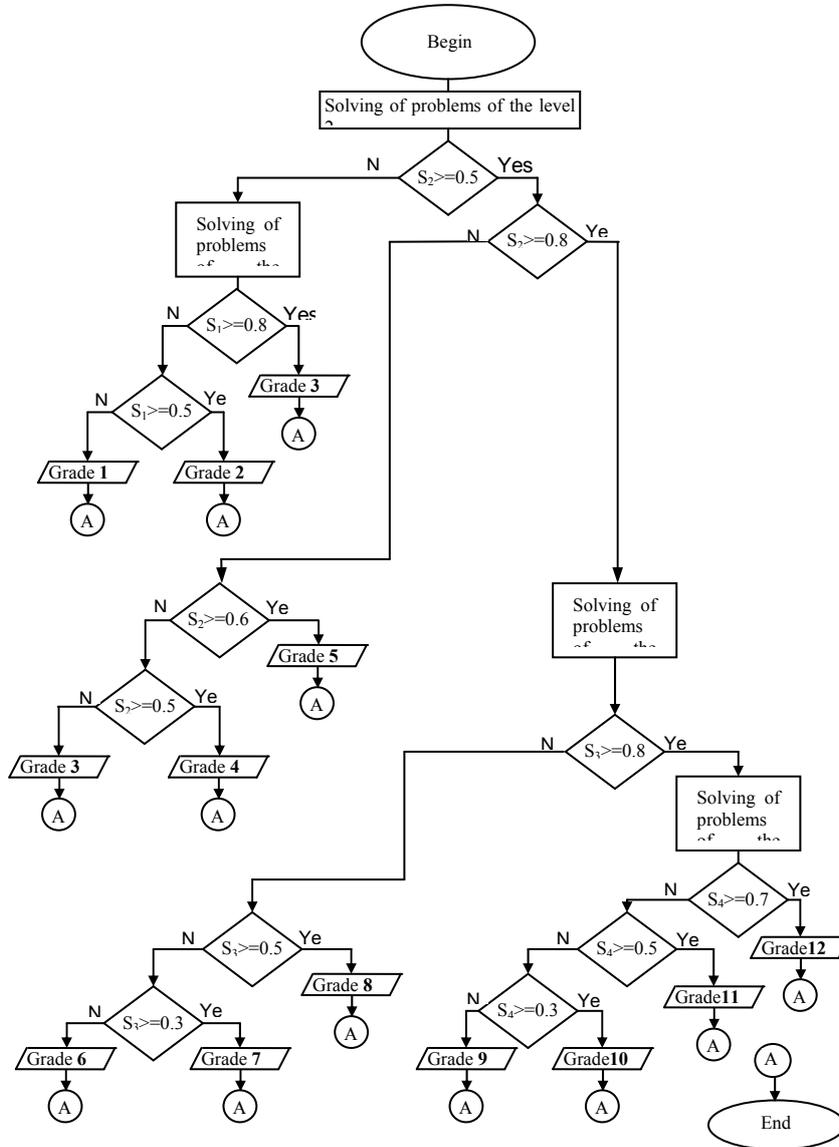


Fig. 2. Grading algorithm

Testing on the levels 3, 4 and 1 is carried out by analogy with level 2 with scores S_3 , S_4 and S_1 accordingly.

Such algorithm of testing improves accordance between the level of items being suggested and the level of the examinee's mastering. Influence of a lack of attention on the test score for examinees with an excellent mastering is decreased. The exami-

nees with bad mastering solve easy problems, which correspond to the most important parts of the educational content. The psychological discomfort, connected with constantly incorrect answers, is excluded, but such easy items give a possibility to determine the structure of the knowledge and skills, as well as to distinguish examinees, who have not mastered the compulsory minimum according to curriculum.

In any case the result of the pedagogical diagnostics will be more careful in comparison to testing without special selection of items.

4 Software

The computer support of offered technology is provided with information system "Expert 3.05" designed by us as a distributed database in the MicrosoftAccess2003 environment. The important advantage of our information system in our opinion is a modular principle of its construction which allows the author of the test items (probably, with the help of the programmer) to create and to add to the database new forms of the test items. The central database of information system is the test items database. The test items are grouped by topics for convenience of viewing. The elements of an educational content are picked out in each topic. To each element the author provides the comment for a student who has not mastered this content. Some blocks of the test items of a different level of difficulty are offered to verify student's mastering in each element of an educational content. The author specifies for the every block such parameters: a test items level (0-4), a weight factor and maximum time of exposition of one item. All items of the block should be of one type, that is, the identical dialogue form. The student will be offered one or several items from each block by a casual choice in a process of testing. Quantity of items blocks and filling these blocks are determined by required quality of diagnostics.

The database, which contains the information on the answers of each examinee on each item, is formed by results of testing. This database includes such fields: the code of the test item, the correctness of the given answer, the level of the item, probability of casual guessing of a correct answer, time of item solving. The additional service information (time and date of testing, examinee's grade etc.) is also being stored.

The examinee receives (fig. 2) the diagnostic data on each element of knowledge; chart, which reconstructs a structure of his (her) knowledge; recommendations for independent work. The author receives the statistical analysis of each item difficulty and discrimination, correlation with the test score (grade). The diagram shows dependence of item difficulty on the examinee's test score (grade) and is very useful (fig. 3). The author has an opportunity to generate database query by means of Access environment and to pass the data for the further analysis in spreadsheets.

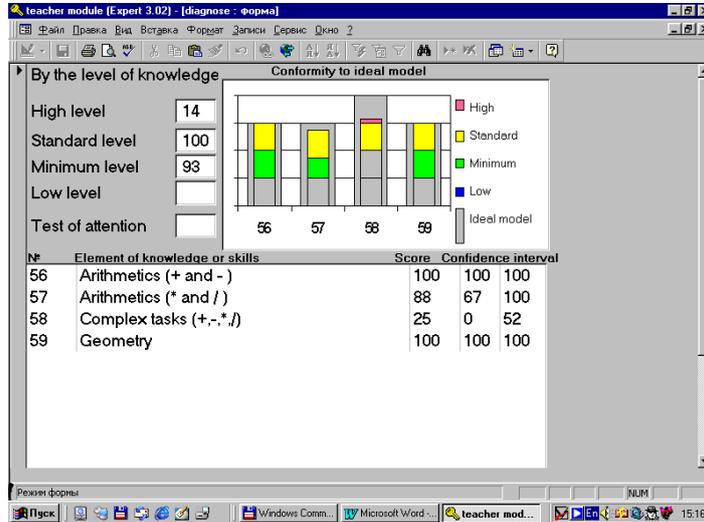


Fig. 3. Diagnostic data for the examinee

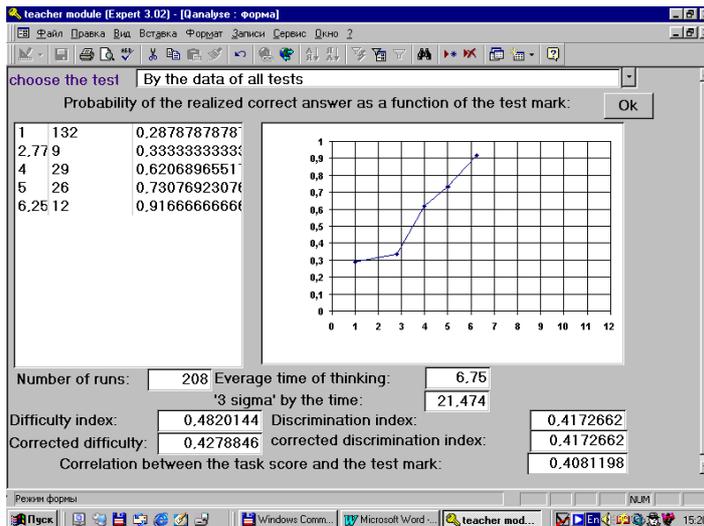


Fig. 4. Statistical analysis of a test item for the author

5 Experience of Diagnostics with “Expert 3.05”

We have prepared the system of the test items with "Expert 3.05" on some courses: “Mathematical methods in psychology”; “Theoretical basis of informatics”; “Architecture of a personal computer”.

Now we are able to start the third stage of the preparation of the system of the pedagogical diagnostics - approbation and verification of the test items.

Our technology of verification is based on the requirements of the Standard of the Ukrainian Ministry of the education and science [6] and takes into account features of the automated pedagogical diagnostics.

The analysis of the test begins with detecting of a level of educational achievements of the students based on the expert's rating, for example, it may be a traditional examination. The complete verification procedure assumes that the experts determine such rating irrespective of the verifying test. Approbation should be organised with enough number of examinees to guarantee sufficient number of answers for any test item. Determination of the student's rating by experts cannot be organised so often, as it is necessary for constant updating of the test items database. Therefore, for the current verification it is possible to offer detecting of a level of student's educational achievements with the help of the same automated system of pedagogical testing and to check a correlation of the separate item with the test score. In such a case, the approbation data are accumulated continuously, including the independent work of the students with the automated system. Validity of the automated system of testing as a whole is checked through comparison of integrated results of testing with results of other kinds of the control: interview, examination, execution of practical works etc. For maintenance of reliability of the current verification, the automated system does not include into the analysis the answers of not registered examinees: teachers and other users, whose names are not in the lists of the students groups. The answers of the test pass, which has not been finished, are not analysed too. The answers are not taken into account, if the time of its execution is smaller, than it is necessary for acquaintance with the text of the problem. There is an opportunity to specify additional conditions of selection of the valid answers with use of the Access environmental (for example, date and time of testing, educational group, variant of the test etc.).

After distribution of the students according to their educational achievements, the conformity of a level of the test item and its empirical index of difficulty is checked. According to the requirements of the Ukrainian educational standards, the students with an average level of educational achievement (the grade of 4 on the grading scale with 12 grades) "... knows about half of educational content, is capable to reproduce it, to repeat after the model the certain operation..." [3]. Just the items of the level 2 in our classification have such contents. Thus, the difficulty index of the items of the level 2 can not be less than 0.5 for such students and we consider it to be within the range 0.5-0.9. It is necessary to note that such a range of difficulty is not convenient from the viewpoint of improvement of test statistical parameters. However, the items of the level 2 represent a set of educational content facts, which are obligatory for mastering. Therefore, the problems author cannot change its difficulty without changing the curriculum.

Thus, for the items of the level 2 on the sample of students with an average level of educational achievements we have the following algorithm of analysis of the item quality:

- An item does not require correction, if its difficulty index within 0.5-0.9, the discrimination index is higher than 0.25, that is, discrimination index satisfies the requirements of the standard [6] (fig. 4.).

- The item difficulty index is more than 0.9. This item should be analysed with the help of the diagram of dependence of difficulty from educational achievements of the students. It should be determined, whether this item has the discrimination ability for the students with the initial level of educational achievements, accordingly, this item level should be changed (fig. 5). Otherwise, this item should be removed from the test.
- The difficulty index is less than 0.5, it is necessary to analyse the content of the item, such situations are possible:
 - The item is not reproducible, its discrimination ability satisfies the requirements of the standard. It is necessary to increase the item level (fig. 6).
 - The item has low discrimination ability for all students; it signifies that some mistakes in the formulation of the item take place. This item should be removed or corrected (fig. 7). There is a possibility of the situation, when all experts agree that the problem is correctly designed and satisfies the curriculum, in such a case mastering of the students should be checked by some another method and, may be, the quality of educational process should be analysed.

The best range of difficulty index for items of levels 1, 3, 4 will be within 0.5-0.6 (allowable 0.3-0.7) on the sample of the examinees of appropriate level of mastering. The analysis of these items on a difficulty is carried out by analogy with level 2.

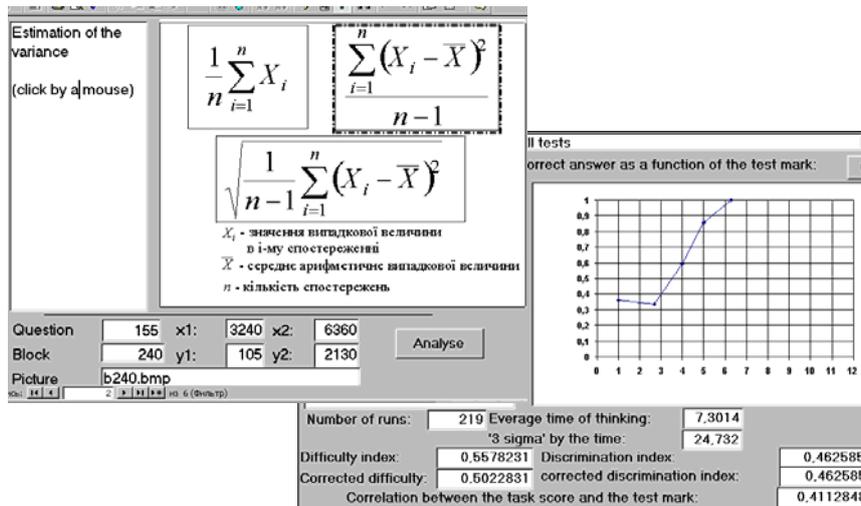


Fig. 5. A typical item of the level 2

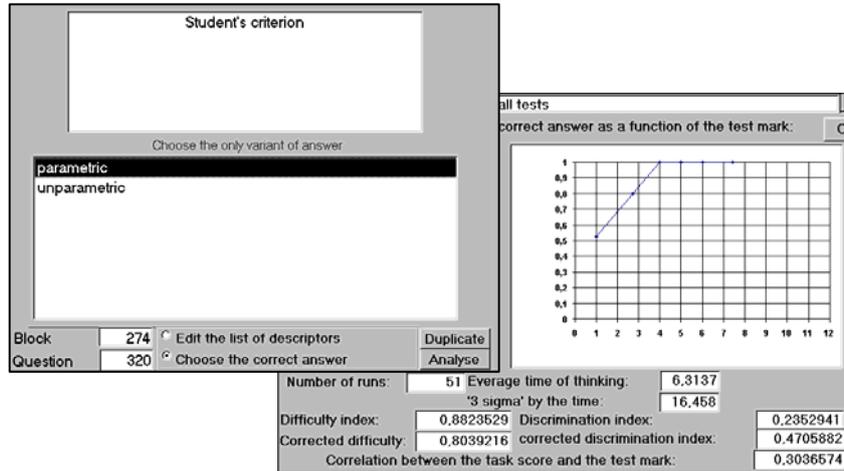


Fig. 6. A typical item of the level 1

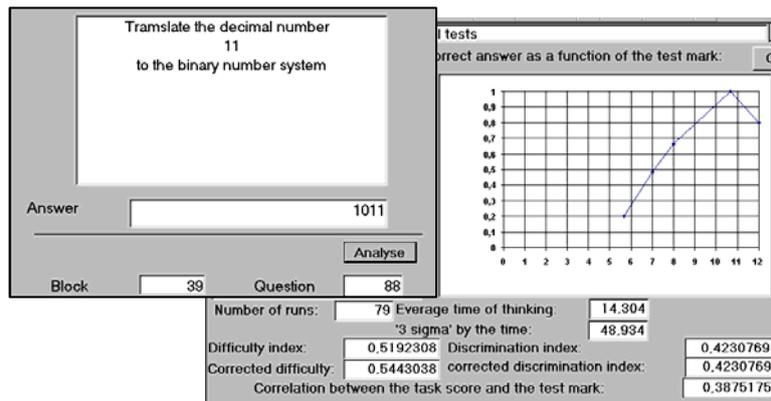


Fig. 7. A typical item of the level 3. It is not a reproductive problem, if the students do not learn the table of the binary codes for numbers until 15.

After finishing three stages of preparation the system of pedagogical diagnostics is ready for practical use. The stage of practical application of the system combines procedures of testing and statistical processing of the obtained results including the interpretation of the results for students, teachers and authors of the test problems. The expert system of pedagogical diagnostics needs continuous modernisation of its database. Naturally, it requires returning to previous stages of the work with the system.

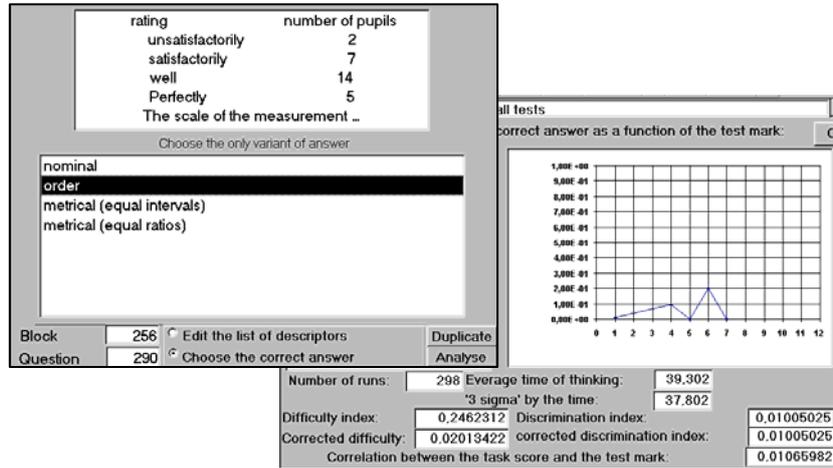


Fig. 8. An unsuccessful item

The “Expert 1.01-3.05” software is used in Kharkiv National Pedagogical University named after G.S.Skovoroda since 2001. Here are some latest results on approbation. The test on mathematical methods of statistical analysis of the pedagogical diagnostics data was suggested to future teachers of informatics, mathematics, chemistry as an element of courseware "Information systems in pedagogical activities" in 2012-2013 academic year. The purpose of testing was the self-diagnostics of students. So, the students could pass the test many times, studying the problem elements of the learning content and improving their results. We took into account the best results of testing and compared its with the examination results. The Pearson correlation was 0.7 at sample of 51 students and we can consider it as the test validity.

The test results gave us possibility to study the structure of students’ knowledge at basic questions of statistical analysis of the pedagogical diagnostics data (fig. 8). The error estimation for the data on fig. 8 was evaluated as a half of the 95% confidence interval

$$\Delta y = \frac{1.95s}{\sqrt{n}},$$

where s is the estimation of the standard deviation and n is the number of test items of the given learning element, which were passed by students. The errors are different for every point on fig. 8, because the number of test items on various elements is different, so we show the ranges of errors in table 1.

The results (fig. 8) show that problems of choosing the scales for the pedagogical evaluations are the most difficult for students. The problems of reproductive level, where student should to choose the method or formula for estimation of some parameters of statistical distribution, are the easiest. But on the productive level, when students should explain the influence of the values and number of variants in a sample on the estimated parameters, such problems are the most difficult.

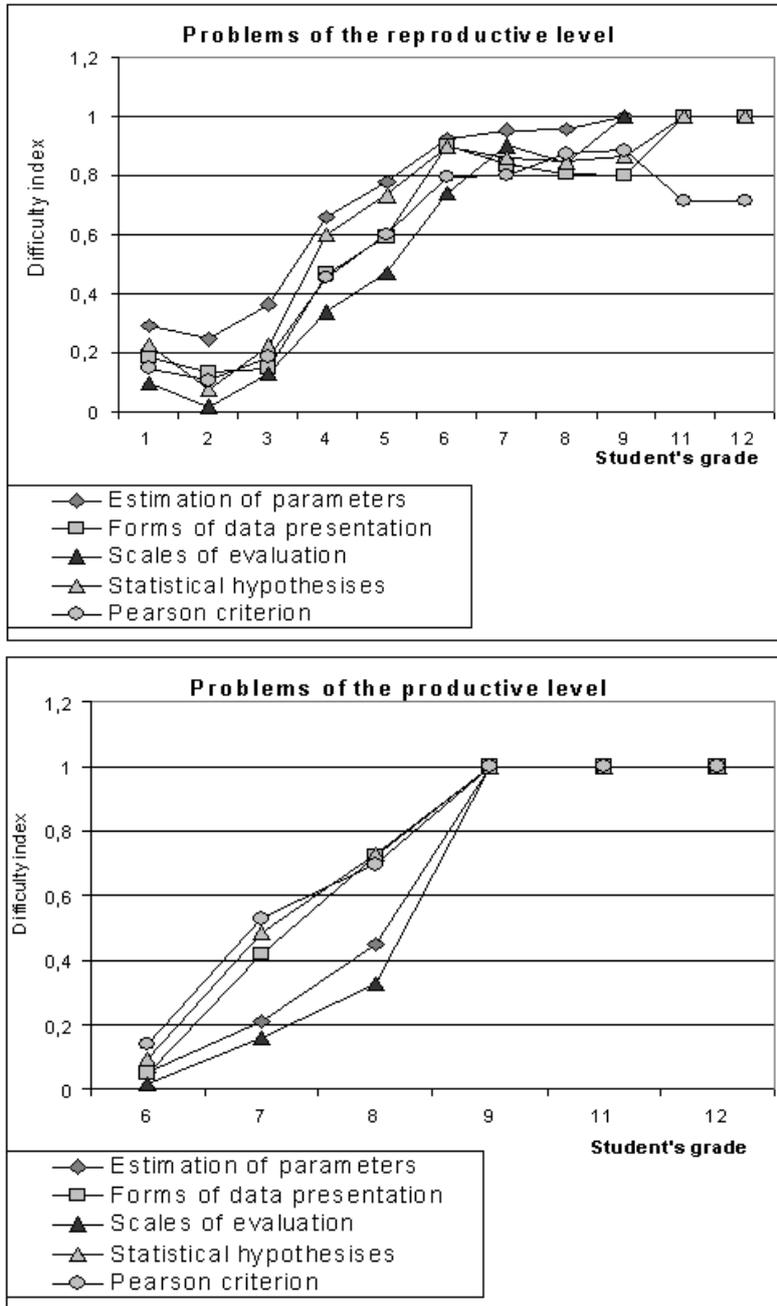


Fig. 8. Difficulty index (probability of correct answer) as a function of the student's grade for problems of various elements of the learning content

Table 5. Errors of difficulty index estimation at different student's grades

Student's grade	Estimation error
1	0.01-0.04
2	0.04-0.1
3	0.02-0.05
4-5	0.01-0.07
6-7	0.02-0.16
8	0.03-0.18
9	0-0.4
10-12	0-0.6

6 Conclusions

1. New comprehensive algorithm of testing and grading is suggested. This algorithm takes into account possibilities of the computer technologies and requirements of Ukrainian standards.
2. The automated system of the pedagogical diagnostics is designed.
3. The methods of the items database administrating is proposed and used in practice in the educational process of the Kharkov National Pedagogical University.

References

1. Bloom, B.S.: Taxonomy of Educational Objectives. Book 1: Cognitive Domain. Longman, Inc., New York (1956)
2. Bepalko, V.P.: Basis of the Pedagogical Systems Theory: Problems and Methods of Psychological and Pedagogical Providing of Technical Teaching Systems, Voronezh University Press, Voronezh (1977)
3. Criteria of Grading of the Educational Achievements of Pupils in the System of Secondary Education. Education of Ukraine, 6 (2001)
4. Zaitseva, L.V.; Prokofyeva, N.O.: Models and Methods of Adaptive Knowledge Diagnostics. Educational Technology & Society, 7, http://ifets.ieee.org/russian/depository/v7_i4/pdf/1.pdf (2003)
5. Kolgatin, O. G.: The Statistical Analysis of the Test with Different Forms of Items. Means of Teaching and Research Work, 20, KhSPU, Kharkiv (2003)
6. Means of Diagnostics of a Level of Educational and Professional Training. The Tests of the Objective Assessment of a Level of Educational and Professional Training. Order of the Ministry of Education and Science of Ukraine, № 285, 31 July 1998 (1998)