Intelligent Testing Systems Based on Adaptive Algorithms

Mykhailo Koliada [1] [0000-0001-6206-4526], Tetyana Bugayova [1] [0000-0003-1926-16], and Nina Miklashevich [2] [0000-0002-0536-8926]

[1] Donetsk National University, Shchorsa Str., 17, 283100 Donetsk, Ukraine
kolyada_mihail@mail.ru, bugaeva_tatyana@mail.ru
[2] Donbass National Academy of Civil Engineering and Architecture, Derzhavina Str., 2, 286123 Makeyevka, Donetsk region, Ukraine
mnv57@mail.ru

Abstract. The paper presents the capabilities and distinctive features of intelligent testing systems based on adaptive algorithms. The list of mechanisms inherent in computer systems for testing the ideas of artificial intelligence is shown. Possibilities of intellectually adaptive testing are revealed by the example of the mechanism of "parallelizing" the system's own operation. Methods of solving the problem of assessing the level of formation of students' functional literacy, traditional input testing, as well as formalizing vague, uncertain and fuzzy test answers have been proposed. The solution of obtaining the simultaneous increase in the measurement efficiency by all criteria is given, in particular, when in adaptive mode the greatest attention is paid to minimizing the test time and the number of tasks presented, at the same time questions of the accuracy of the marks are fading into the background. The review of the criteria selection pattern based on the so called" constructivist approach" has been suggested. It was established that in adaptive testing it is possible to optimize the correlation of the difficulty of the tasks themselves and their number, which had led to the emasculation of the content of the test .The method of measurement objective errors correction while testing diverse characteristics of the students by alignment of different measurement scales has been substantiated. On the basis of the "floating optimization approach" the solution of the problem associated with the assessment of those students who have done the tasks with the maximum number of matches , compared to those who have solved more difficult tasks but with less number of matches has been presented. It is proved that the invariant calibration of tasks of the adaptive system improves not only the semantic quality of new tasks, but also enables to introduce modern innovative forms of their presentation. A new mechanism for controlling the accuracy of entering students' answers based on the Michael Damm algorithm has been suggested.

Keywords: Intellectually Adaptive Testing, Assessment Criterion, Measuring Scale, Invariant Calibration of Tasks, Input Accuracy Control, Damm algorithm.

Proceedings of the 4th International Conference on Informatization of Education and E-learning Methodology: Digital Technologies in Education (IEELM-DTE 2020), Krasnoyarsk, Russia, October 6-9, 2020.

1 Introduction

The diagnostics and knowledge control systems of students hold a special place among the contemporary intelligent learning systems. The main requirements imposed on the new learning systems include, above all, *intelligence, scalability, openness, flexibility and adaptability* at all stages of the learning process [1].

Even in the process of using machineless testing, the creators of the tests noticed that when testing students, both too easy and too difficult tasks, as a rule, do not provide any benefit for learning. When doing simple tasks, the students waste time without any learning effect, and doing complex tasks, they also idle the allocated time. These tasks become simply useless, because in the first case they make the students do the simplest unnecessary tasks, and in the second case they put them in the position of "the impossible to do dead lock". If the student is well grounded, then it makes no sense to suggest very easy (simple) tasks, and on the other hand, if he, on the contrary, is a low-performing student, then there is a high degree of probability that he will not do very difficult (complex) tasks properly (or in general, he will stop doing them). Therefore, we face the problem to avoid these extremes and try to select the testing tasks according to the corresponding level of students' competence.

Historically, when there was still no mainstream use of computer technology it was very difficult to solve this problem, but the appearance of computers has made this task quite feasible. An automated computer testing system began to adapt to the learner: when the test task is successfully completed, it offers the learner the next more difficult task, and if the learner fails to do it, the system gives easier task. Therefore, adaptive test is the "*test in which the complexity of the subsequent tasks depends on the correct answers to previous tasks: the more correct answers to previous tasks, the more difficult the subsequent ones*" [2, p. 42]. But if a testee cannot complete the task, then the automated system reduces the level of complexity and offers him a new easier task. M. B. Chelyshkova defines adaptive testing as "*a set of processes for generating, presenting and evaluating the results of fulfilling adaptive tests, which provides an increase in measurement efficiency compared to traditional testing by optimizing the selection of task characteristics, their quantity, sequence and presentation speed as applied to the training characteristics of tested students*" [3, p. 28]. In such testing there is a constant adjustment of the task according to the difficulty to the level of students' competence: the next tasks are selected by "tuning" for the current answers (estimates), which vary depending on the results of each previous test task.

The adaptive testing has great advantages when it is based on artificial intelligence facilities.

2 Problem History

The first testing in a contemporary implementation was carried out by J. Fischer in the UK in1864 to verify the level of learners' competence using the original special books (scale books). Theoretical fundamentals of testing were developed later, only in 1883 by the English psychologist F. Galton in his work "The study of human abilities and their development." The term "test" was first introduced by the American psychologists J. Cattell and B. McKeon in their book "Mental tests and measurements" in 1890 [4, p. 118].

The first ideas of adaptive testing appeared in the early 60s of the 20th century, first based on the ideas of the classical theory of tests and then on the so-called "modern test theory» (Item Response Theory – the IRT). It was in the 80s of the 20th century when the theoretical and technological basis of contemporary methods for generating adaptive tests was laid. A. Anastasi [5], F.B. Baker [17], B. D. Wright [7], D. J. Weiss [8] and especially F.M. Lord [9] made a great contribution to the development of scientific understanding of the new type of testing. They have carried out a large-scale research activity on adaptive testing based on IRT scientific apparatus within the framework of the Educational Testing Service program (ETS).

The scientists like M. J. Ree and H. E. Jensen [10], H. Swaminathan and J. Gifford [11], R. K. Hambleton, J. N. Zaal and H. J. Pieters [12], W. A. Sands [13], A. R. Zara [14], R. R. Meijer and M. L. Nerling [15] were engaged in adaptive testing research implementing it in a computerized environment .Theoretical problems of using computer-based testing evolved in the works of W. Linden and A. W. Glas [16], E. C. Papanastasiou [17], S. Laumer, A. Stetten, A. Eckhardt [18]. The practical side of the automated adaptive testing was considered by C. G. Parshall, J. A. Spray, J. C. Kalohn and T. Davey [19], N. Wainer [20]. J. Piton-Gonçalves and S. M. Aluísio [21; 22], L. Kroker [23] and others studied the multidimensional computer adaptive testing in the information and educational environment.

Russian researchers V. S. Avanesov [24], V. A. Wexler [25; 26], L. I. Gerasimova, E. V. Gerasimova [27], V. I. Zvonnikov [28], N. F. Yefremova [29], S. N. Larin [27], A. N. Mayorov [30], N. T. Minko [31], O. Yu. Nikiforov [11], V. A. Otrokov [26], M. B. Chelyshkova [28; 3] and others considered the use of adaptive tests from a pedagogical point of view. But the problem of using precisely intelligent adaptive systems in the educational process is insufficiently investigated by domestic and foreign researchers.

3 Materials and Methods

The first appeared forms of adaptive testing have already had the simplest elements of intelligent analysis, but it would be wrong to call them *intelligent testing systems*. Test models based on the ideas of artificial intelligence in addition to the implementation *of adaptive algorithms for testing*, should still have a whole range of opportunities inherent in intelligent information systems. They include the following:

– the mechanism for *making optimal decisions* (including the mechanism of “what will happen if ...?);

– the mechanism of *special function capabilities* (classifying patterns, clustering objects – correlating elements into groups, approximation of functions , etc.);

– the mechanism for *formalizing and interpreting the judgments of the testee* (using the theory of fuzzy sets and the theory of fuzzy logic);

– analyzer for *revealing the complexity of test tasks*;

– the mechanism for *identifying the meaning of the text*;

– the mechanism for *searching own patterns and rules* (Data Mining);

– the mechanism for *considering internal and external factors affecting the quality test answers*;

– the mechanism of *correction* ("adjustment") *of the testing system in accordance with the individual and psychological-typological characteristics of students*;

– the mechanism for *developing one's own criteria for evaluating test tasks*;

– the determinant of *correlations and relationships between tasks and groups of test tasks*;

– the mechanism of "*parallelization*" *of one's own operation in several directions*.

It is quite an incomplete list of mechanisms and tools involved in functioning of the system of intelligent adaptive testing.

Considering the *mechanism of "parallelizing" the operation* of the intellectual adaptive system in several directions of its functioning, we often face the problem of simultaneous (parallel) research of hidden (*latent*) features of students, namely the *identification of the flexibility of thinking; rate of analysis, generalization and synthesis; level of analytical and synthetic activity* and many other individual properties and personality characteristics. This happens when several mechanisms tracking indicators of different nature and direction are activated simultaneously. The results of this programmed intellectual- functional operation of the system are entered (stored) in the data bank, and then they are used by other diverse mechanisms for analyzing latent characteristics of the students. The peculiarity of this double detection is characterized by the fact that new objective information obtained in this way is usually not available to be found using other ways, particularly those associated with the study and evaluation of typological properties and psychological characteristics of the tested individual.

Intelligent testing systems cope successfully with the problems of *assessing the level of formation of students' functional literacy* [3, p. 157]. For this purpose the fragments of texts in which *errors are known to be present* are used. The students under testing are asked to correct them by rewriting the pieces (sections) of this text. Sometimes the tasks based on the constructed text ans wers are suggested. The student under testing has to compose a short essay or micro essay on a given topic. Assessment of such tasks is possible only in programs based on the ideas of artificial intelligence , since the criteria for their assessment are quite complex linguistic characteristics, such as the *quality* (stylistic, grammatical, etc.) and *clarity of presentation; length, degree of completeness of the answer; level of imaginary narration; the degree of covering the topic*, etc. It's impossible to do without the use of the mechanism for revealing *the meaning of the text and the analyzer of text information complexity*. The system of automated assessment of the essays will involve mechanisms based on the achievements of computer linguistics [2] and make proper assessment of these tasks.

It should also be pointed out that in computer testing systems the necessity for two-step adaptive testing arises. It is necessary to pass the so-called *intake testing*, and only then the adaptive one. But unlike the traditional approach, the self-learning mechanism of the intellectual system based on the ideas of the operation of artificial neural networks is used here. The computer program of artificial intelligence takes a huge set of initial input data with many variables, in which the patterns are not yet known. Then it analyzes these data, processes the relationships between them (in the form of correlations), and only then it selects a set of those variables that are similar to the reference value (models). It is this initial testing that becomes the starting point for evaluating subsequent adaptive testing results. Based on this preliminary conclusion, the program changes models by adjusting the parameters of variables or even, if necessary, excludes them from analysis and evaluation. It repeats this procedure many times, each time improving its previous model (and result), the best options being stored. If during such iterations the further improvement of the model does not occur, it pauses its work and displays the best result as the final one. In a similar way, an intelligent adaptive *system self-learns* in other areas of its functioning.

Unlike classical testing methods, intelligent systems can assess the degree of correctness of the students' answers fairly accurately even when the wording of the students' answers is very vague and ambiguous. When analyzing the answer in a test task, a person tends to use statements consisting of key inaccurate words such as "almost", "a little", "approximately", "like", and the like. Using the mechanisms of fuzzy logics, the *system formalizes these answers*, processes them based on clear mathematical rules of the theory of fuzzy sets, and it displays very objective results of scoring [32, p. 211].

As a rule, in classical adaptive testing it is impossible to obtain a simultaneous increase in the measurement efficiency by all criteria, therefore providing that, one or two criteria come to the fore. For example, in some cases of express diagnostics in adaptive mode the greatest attention is paid to *minimizing the test time and the number of tasks presented*, at this, *questions of the accuracy of estimates* fade into the background. In other cases the accuracy of measurement may be a priority, and testing of each student will continue until the *planned minimum measurement error* is reached [28, p. 163]. In intelligent adaptive systems these difficulties are successfully overcome. Here the so-called *constructivist approach is used* which finds agreement between the indicated contradictions in the form of quantities and conclusions obtained on the basis of our own analysis of the intellectual system. This approach forces us to reconsider the criterion selection scheme itself, adapting the obtained conclusion to its continuous change (improvement), and therefore search models and models for evaluating the result jointly evolve toward "equilibrium improvement" [33].

In classical adaptive testing systems the degree of difficulty of tasks usually reduces their number for presentation, the *content of the test being emasculated* (i. e., the coverage of the entire studied material decreases). Therefore, the *content validity* of the generated adaptive test is not presented. In intelligent adaptive testing the system itself monitors the fulfillment of the norm of the degree of complexity of tasks and their mandatory number for presentation on each subject (section). In addition, it checks the condition of the frequency of selecting tasks from the data bank, and after each completed task the system constantly checks the difference between the received and planned measurement accuracy. Only after reaching the established measurement accuracy it can stop the testing process.

According to the fact that adaptive testing obviously uses tasks of different complexity levels, which are implemented by different methods, types, forms and approaches, different measuring scales are also used to evaluate them. Researchers distinguish four types of scales: *nominal* (scale of names, categorical scale), *ordinal* (rank), *intervals, relations* [9, p.11]. According to the degree of increase of power, the scales are arranged in the following sequence: *names, rank, interval, relations*. From this it can be seen that *non-metric* scales have less power in comparison with quantitative scales, since they contain less information about differences between objects. When calculating test scores, *objective measurement errors occur*, because the tasks of diverse complexity have different forms of entering answers, and they require different measuring scales. It is impossible to correlate the results of answers on changing scales manually, but the intelligent system can easily cope with this difficulty by *aligning the various measuring scales*. To do this, it uses appropriate mechanisms, for example, the method of analyzing hierarchies in decision making.

As you know, the length of the adaptive test is significantly affected by the quality of the structure of students' knowledge [28, p. 163]. Typically, a testee with a clear knowledge structure complete tasks of increasing difficulty, elaborating the assessment of competence with every correctly completed task. They perform a small number of adaptive test tasks and quickly reach the *threshold of their competence* (i.e., to a predetermined level). Testees with a fuzzy structure of knowledge, who alternate between correct and incorrect answers, receive tasks with different range of difficulty. This seemingly negative side of adaptive testing finds a positive solution in intelligent systems. The intellectual system is reconfigured and presents to those who are tested special tasks which form a more *structured knowledge system* for students. But this problem has not yet been fully resolved, scientists are still working on it.

The distance learning system (DLS) standard developed by IMS (Global Learning Consortium, Inc. – IMS) [34] provides three types of adaptive testing:

– *pyramidal* – at first, every testee is given tasks of average difficulty, and then, if the answer is correct, the next (more difficult) task is offered in terms of difficulty, and if the answer is not correct, the previous (easier) task is given;

– *flexible* (flexilevel – English “flexible level”) – the testee chooses the desired (for example, obviously overstated) level of difficulty of tasks [35; 36], and in the case of its successful completion, the transition proceeds upward on the scale of difficulty, and in the case of failure of his answer, downward; such leaps will continue until the real level of the student’s capabilities is established;

– *stratified* (stradaptive, from the English “stratified adaptive” – “sorted”, “selected by criterion”) - testing is carried out on the base of the task selection from the data bank, grouped by difficulty level, and if the answer is correct, the next task is taken from the group of tasks of a higher (complex) level, if not, from the lower (easier) level [1].

As you can see, all three types of adaptive testing, unlike the classical tests, are dynamic that is, constantly changing and on-line. They are not static, where the list of questions proposed to the tested person does not change and is predetermined [1].

Since the procedure of student scoring in adaptive testing is a value dependent on their answers at each step of the test tasks, so it requires the use of the so-called *polytomic evaluations* (i.e., evaluations associated with establishing the number of correctly established correspondences).

In the case of polytomic evaluations, the best ones can be those who performed the tasks correctly with the *maximum number of correspondences*, while other testees who *know more* and completed more difficult tasks with fewer correspondences will be in the worst position and will receive a lower score [28, p. 143]. To correct this drawback, the so-called "*floating optimization approach*" can be used in intelligent systems. This approach is based on the fact that more difficult tasks have a *higher weight score*, which the system necessarily takes into account when calculating the final results.

A well-known testing researcher V.S. Avanesov identifies four *forms of test tasks* [24]:

1. *Choice tasks*, which are divided into 3 subgroups: tasks with a choice of one correct answer or univariate tasks, tasks with a choice of several correct answers or multivariate tasks, tasks with a choice of the most correct answer.
2. *Open tasks*.
3. *Tasks to establish the correspondence*.
4. *Tasks to establish the correct sequence*.

In intelligent systems, the algorithm for varying the choice of questions in test tasks is easily implemented. Thus, when the questions are used again, the disadvantage associated with remembering the number of the correct answer, is easily eliminated. The same disadvantage is avoided when using the same test tasks by other participants (subsequent test persons) who are able to use the correct answers taken from those who have already passed this stage of testing.

In general, the problem of re-presenting the same tasks with pre-known answers for the testees in intelligent systems is also solved at a more global level. To do this, the *task calibration* methods are used [28, p. 160]. A key factor is that the number of tasks put in the database must be *large enough* (so large that when using the test repeatedly, these tasks are not likely to be repeated). One of the main *conditions for calibration* is that tasks in complexity groups should be *invariant*. This feature allows to evaluate such tasks on the basis of even one group of testees, and then confidently distribute (use) them for any other group of people. Under this approach of calibrating tasks you can gradually add new tasks to an existing bank of placed tasks by presenting them to new groups of testees in the form of new test forms containing some of the already tested tasks (which are called the *anchor part*). Thus in future, the intelligent adaptive system enables to present *different forms of the test itself* for different groups of testees using a *single latent characteristics scale*. Such an invariant task calibration is a great value of the adaptive system, since its implementation improves not only the semantic quality of new tasks, but it also introduces contemporary innovative presentation forms. For example, you can use the new multimedia capabilities of the computer: audio, visual (photo and video) series, new interactive forms and approaches for submitting questions, etc. This is something that without losses, but at a *high level of multidimensionality* increases the learning effect and the evaluative qualities of tasks. The multidimensional nature of multimedia technologies leads to the fact that the completion of such a test enhances the functional efficiency associated with their *clarity, interactivity, dynamism* and other positive features that contribute primarily to the development of creativity and non-standard thinking. It should not be forgotten that excessive oversaturation with sound and visual images in computer testing often distracts the testee, confuses him from the main idea, and younger schoolchildren feel tired. Therefore, an intelligent system must necessarily monitor the optimal number of multimedia-formed tasks, using a control module that compares the psychological-typological and personality-physiological characteristics of the data of a particular tested person and shows an approaching threshold of fatigue (or other danger), or switches on automatic re-profiling of the system to reduce or isolate tasks leading to negative phenomena.

The main feature of the presentation of questions of the adaptive test is the *variation of their difficulty level*. But if in the traditional use of adaptive testing the selection of such tasks was carried out through their preliminary empirical selection and approbation of typical students on a sufficiently large sample (as they say, taking into account the general population) [30], then in intelligent adaptive testing this process is replaced by the selection of tasks according to the difficulty groups simultaneously with it (and sometimes independently on it) on the base of specially developed artificial intelligence algorithms. They include such algorithms as *ant colony optimization*, *evolutionary* (C. Darwin's algorithm), *immune system algorithm*, *annealing algorithm*, etc., although the models for the optimal *classification* and *clustering* testing tasks are still the main instruments.

4 Research Results

It is commonly known that the control of the test accuracy is carried out by comparing the answer of the tested person with the standard answer of the test, more precisely, by identifying the entire list of distracters (plausible, but incorrect answers) and accordingly, choosing the correct answer.

But in adaptive testing implemented in machine execution, a computer randomly selects questions of the same difficulty level from the database of tasks. It can also implement a variation when among the proposed answers to the task, their sequence is also generated in a stochastic way. When proceeding to questions of another (more difficult or easier) level of tasks, the computer again randomly generates questions for presentation to the tested person with possible mixing (probabilistic generation) of the proposed answer sequence. In this case, *controlling the verification of the correct / incorrect answer* is a major problem. The mode of determining the testee's answer usually comes down to registering the number of the task itself (question) and the corresponding number (code) of the correct answer, but since the numbers of the distracters and the correct answer are constantly changing (more precisely, their sequence for presentation), it is very difficult to establish the true correspondence of the task assessment and reference value. In this case, a *way of coding the designations* of the question numbers (tasks) and numbers of distracters (together with the correct answer) comes to the rescue.

To link the number of the question and its correct answer (that is, the establishment of the so-called "label"), there are many invented techniques that have been used in cryptography. Here is one of them, based on a probabilistic basis.

They take some four-digit number (suppose 41 3 2 is the threshold designation of the three distracters and the correct answer - only four positions) with the correct answer under number 3 (this is our mark in the third position), they take the square of it (17073424), and the average figures (0734) are written out of the result obtained (17 0734 24), considering them as random ones. Now these “random” numbers are squared, and the centre of range is again extracted from the result obtained, etc.

They stop this procedure for output only when it satisfies some established rule (condition), for example, when these average digits will stand next to each other for the first time in repeating numbers from 1 to 4 in any sequence (four is the number of suggested answers). It is also possible that such a quadruple that does not repeat in a row appears in any other place of this complete sequence, or in accordance with some other intricate restrictive algorithm. In this case, the entire chain of executed actions is registered (stored) by the intellectual system for repetition, if necessary. The sequence obtained with the help of this algorithm can be generally considered random (in fact, it is pseudo random), and the correct answer will be in the position where figure3 is marked by stochastic way (we have this label). Suppose that for the second question (of the same level of difficulty) it will be some obtained position of the distracters and the correct answer, say 241 3, and for the third task - 3 124, etc.

Then the encoded set of correct answers of the test for the first nine questions may have the following form:

1st question: 41 3 2 → 3;

2nd question: 241 3 → 4 (four means the fourth position, which, with random generation, is the number 3 - that is, the label of the correct answer);

3rd question: 3 124 → 1;

4th question: 214 3 → 4;

5th question: 1 3 24 → 2;

6th question: 3 411 → 1;

7th question: 142 3 → 4;

8th question: 421 3 → 4;

9th question: 1 3 24 → 2.

If the system implements questions (tasks) at other levels of complexity, then the picture will not change, and will be similar to this:

3, 4, 1, 4, 2, 1, 4, 4, 2.

Those who are tested have to enter this set of numbers. If the testee confuses at least one number, then his objective assessment will be doubtful. How to avoid such errors? The answer is simple: you must be extremely careful. But when the student has to answer the test questions for a long time and what is more, the test changes its difficulty according to his answers, than “misprints” are sure to occur. In this regard, the ways to avoid these errors have been invented.

In information security systems, this approach is called *hashing*, and its essence lies in the fact that one more, the so-called *check digit*, is added to this sequence of digits at the end (or at the beginning). This digit is obtained by using a special algorithm based on the positions of the introduced correct answers codes, and if the testee confuses one of the digits in the sequence of digits that he enters, the intelligent testing system having calculated the check digit from the codes already entered, immediately detects this error.

There are several algorithms for finding this check digit. The first of them is well known enough, and it is called the *modulo addition* method. Its essence is that if adding numbers, the computer system receives the result of more than 10 (or several times ten), it only leaves the last digit of the result, and if less than ten, then the resulting digit itself is exposed as if it deals with the usual addition. For example, adding up a series of numbers of the testees' answers obtained only by the first position of the correct answers, we get an amount equal to 9:

$$1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 9,$$

and the system will give the same check digit 9. And in our case, the system considers the amount that is equal to 25:

$$3 + 4 + 1 + 4 + 2 + 1 + 4 + 4 + 2 = 25,$$

then the intelligent system, as a check digit, will only display number 5.

The algorithm developed in 1954 by an IBM employee Hans Piter Luhn allows us to judge about the absence of errors in the block of digits of the testee's answer only with a certain degree of certainty [37]. According to the algorithm of P. Luhn, this check digit is added to the existing nine-digit series of numbers at its end, and a ten-digit number is already obtained, in our case it is: 3, 4, 1, 4, 2, 1, 4, 4, 2, 5.

After entering a set of the correct answers of a testee, the intelligent system will add the resulting series of numbers by 10 modulo and compare the result with this check digit. If the answer does not match, then the system will indicate this, that is, it will immediately detect the incorrect input of the code (numbers) of the correct answers.

For example, if the testee confuses only the number of one answer and enters another one instead, the Luhn's method will definitely detect this error, since the check digit will turn out to be different. But unfortunately, this method does not detect all the errors. For example, if the testee changes the position of two adjacent digits unintentionally, then this method will not find this error. For example, if he enters 4, 3, 1, 4, 2, 1, 4, 4, 2, 5 instead of the reference set of response codes for the test answers: 3, 4, 1, 4, 2, 1, 4, 4, 2, 5, then the system will evaluate it as a positive (correct) result, since the check digit will not change, but this is not so!

Therefore, in 1969 the Dutch mathematician and sculptor Jacobus Verhoeff proposed another way to search for errors based on the values of numbers in a specially selected table (it is called a group) [38]. As before, its algorithm also works through a check digit at the end of the entered series of digits. Such algorithm allows intelligent system to recognize a larger number of errors: first, it finds all the replacements of any digit with the others (as the algorithm of P. Luhn), and it is able to determine 60 out of 90 of all possible replacements of adjacent digits (the number 90 is taken from considerations that he used a table of 10 x 10 digits (9 digits, plus 0 – position row (column) for intermediate initialization).

But the revolution in solving the problem of checking the correctness of the input digital data was made by a German mathematician *Michael Damm* in 2004. In his thesis research [39] he presented examples of the so-called completely *antisymmetric quasigroups* (it was previously believed that such quasigroups did not exist). His algorithm detects all the errors *when the replacement of one digit with another takes place* and it is also able to detect all *single permutations of two adjacent digits*. This algorithm is noticeably simpler and more reliable than J. Verhoeff algorithm comparable in capabilities.

5 Discussion

In the algorithm of M. Damm a way to calculate the check digit was found. It allows to recognize all single replacements of one digit by another one and all permutations of adjacent digits as well.

Suppose we need to find a check digit for the same sequence of test answers 3, 4, 1, 4, 2, 1, 4, 4, 2, but already on the basis of the Damm's table (antisymmetric quasigroup) (Fig. 1). To do this, following algorithm must be executed:

1. At first, the interim digit is initialized to 0.
2. The first number at the intersection of the first code 3 is found, i.e, it's the crossing of the column with the notation "3" and the row, with the notation "0" (initialized interim value). The result returns the number 7.
3. The result of intersection between the second code 4 (column index 4) and the row index with the new value of the interim digit 7 is the number 3. This value is assigned to the new interim digit 3.
4. The result of intersection between the next code 1 (column index 1) and the row index with the number value of the interim digit 3 is the number 7. This value is assigned to the next interim digit 7.
5. The result of intersection between the next code 4 (column index 3) and the row index with the number of the interim digit 7 is the number 3. This value is assigned to the next interim digit 3.
6. The result of intersection between the next code 2 (column index 2) and the row index with the number of the interim digit 3 is the number 5. This value is assigned to the next interim digit 5.
7. The result of intersection between the next code 1 (column index 1) and the row index with the same number of the interim digit 5 is the number 6. This value is assigned to the next interim digit 6.
8. The result of intersection between the next code 4 (column index 4) and the row index with the number of the interim digit 6 is the number 7. This value is assigned to the next interim digit 7.
9. The result of intersection between the next code 4 (column index 4) and the row index with the number of the interim digit 7 is the number 3. This value is assigned to the next interim digit 3.
10. The result of intersection between the next code 2 (column index 2) and the row index with the number of the interim digit 3 is the number 5. This value is assigned to the next interim digit 5.
11. The input sequence is over. The last value of the interim digit (5) is the check digit. The input sequence with the addition of the check digit to the end of it is: 3, 4, 1, 4, 2, 1, 4, 4, 2, 5.

The check digit algorithm for the Damm system is exactly the same as it was described above. However, the 11th step will be changed.

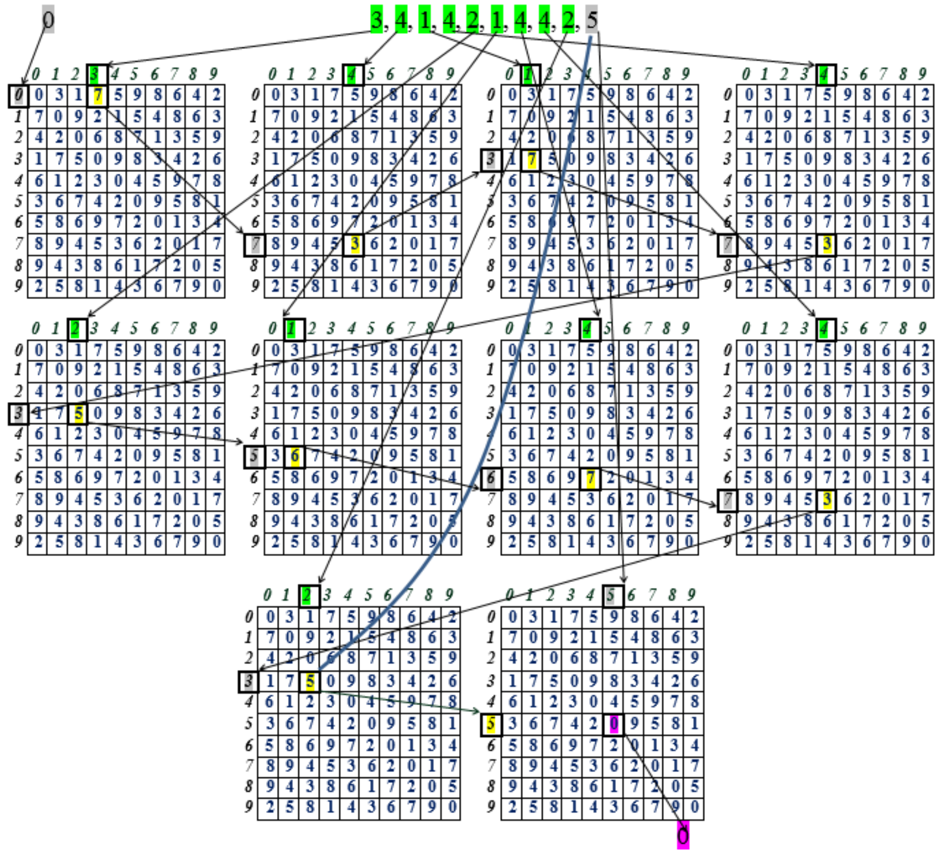

Fig. 1. The scheme of the Damm algorithm for controlling the accuracy of entering codes of the correct test answers

11. The result of intersection between column index 5 (this is the last digit of the input sequence) and row index 5 (this is the value of the interim digit) (see fig. 1) is 0. This value is assigned to the intermediate digit.

12. The input sequence is over. The last value of the interim digit is 0, as it was expected.

Now, if the result of intersection between the check digit and the last interim digit is *zer*o (the diagonal of the Damm table), then it indicates that all the answers to the questions (tasks) of the test are *entered correctly* and there are no errors. Otherwise, the intelligent system will ask to retype the answers correctly.

6 Conclusion

In summary, an intelligent adaptive testing system is distinguishable from classical control methods by a greater objectivity, efficiency, a higher level of differentiating and individualization, comparability of results and a higher degree of intellectualization. All these characteristics involve testologists and practitioners to differentiate test tasks according to the level of subjects' studying.

Having determined the current level of implementation of educational achievements, teachers use this innovative type of testing receiving reliable and timely information on the educational process, more precisely, on how effective and productive the pedagogical system is. Another important aspect of the intellectual-adaptive test control is an effective feedback, as well as the ability to measure the dynamics of achievements and the degree of development of each student, as well as consolidate their knowledge and skills [12]. The indisputable advantages of adaptive testing technology are also the objectivity of assessing the level of this knowledge and skills, their comparability and the possibility of real verification, which are ensured by precise “adjustment” to the typological characteristics of the student’s personality [40, p. 48].

The offered algorithm for checking the correctness (accuracy) of the input of students' answers based on M. Damm's quasigroups is a reliable protection tool against accidental misprints (errors) of testees, as well as an important component of the control of a powerful objective assessment mechanism in an intelligent adaptive testing system. The toolkit of technological intellectually filled opportunities of adaptive testing contributes to the efficient usage of time in the classroom and increasing the productivity of the entire educational process, as well as practical introduction of individualization and differentiation approaches of education.

References

1. Nikiforov, O. Yu.: Ispolzovanie adaptivnykh sistem kompyuternogo testirovaniya [The Use of Adaptive Systems of Computer Testing]. Gumanitarnye nauchnye issledovaniya. 2014. no 4. Available at: URL : http://human.snauka.ru/2014/04/6274 (In Russian)
2. Shermis, M. D., Burstein, J. C.: Automated Essay Scoring : A Cross-Disciplinary Perspective / Mark D. Shermis and Jill C. Burstein (editors), Florida International University and ETS Technologies, Inc. Mahwah, NJ : Lawrence Erlbaum Associates, XVI, 2003. 238 p.
3. Chelyshkova, M. B.: Adaptivnoe testirovanie v obrazovanii (teoriya, metodologiya, tehnologiya) [Adaptive Testing in Education (theory, methodology, technology)]. M. : Issledovatelskii tsentr problem kachestva podgotovki spetsialistov, 2001. 165 p. (In Russian)
4. Kovalchuk, J.O.: Teoriya osvitnikh vymiriuvan’ [The Theory of Educational Measurements]. Nizhin : Vidavets PP Lisenko M.M., 2012. 200 p. (In Ukrain)
5. Anastasi, A.: Psychological Testing (3rd Ed.). New York : Macmillan. 1968.
6. Baker, F.B.: The Basics of Item Response Theory. Portsmouth NH : Heinemann Educational Books, 1985. 131 p.
7. Wright, B. D., Stone, M. H.: Best Test Design. Chicago, MESA PRESS, 1979. 222 p.
8. Weiss, D. J. (Ed.): New Horizons in testing. N. Y. : Academic Press, 1983. 344 p.
9. Lord, F. M.: Applications of item response theory to practical testing problems. Hillsdale, NJ : Lawrence Erlbaum associates. 1980.

10. Ree, M. J., Jensen, H. E.: Effects of sample size on linear equating of item characteristic curve parameters. In D. J. Weiss (Ed.). Proceedings of the 1979 computerized adaptive testing conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory. 1980.

11. Swaminathan, H., Gifford, J.: Estimation of parameters in latent trait models. In D. Weiss (Ed.). Proceedings of the 1979 computerized adaptive testing conference. Minneapolis : University of Minnesota. 1979.

12. Hambleton, R. K., Zaal, J. N., Pieters, H. J.: Computerized adaptive testing: theory, applications and standards // Kluwer Academic Publishers. Boston, MA, US, 1991. 458 p.

13. Sands William, A. (Ed), Waters Brian, K. (Ed), McBride James, R. (Ed), Washington, DC, US : American Psychological Association. xvii, 1997. 292 p. DOI: 10.1037/10244-000

14. Zara Anthony, R.: Using Computerized Adaptive Testing to Evaluate Nurse Competence for Licensure: Some History and Forward Look, Advances in Health Sciences Education, V. 4, no 1, 1999. P. 39–48. DOI: 10.1023/A:1009866321381

15. Meijer, R. R., Nering, M. L. Computerized Adaptive Testing : Overview and Introduction, Applied Psychological Measurement, 23 (3), 1999. P. 187–194.

16. Computerized Adaptive Testing : Theory and Practice / Ed. by Wim J. van der Linden and Cees A. W. Glas. London : Kluwer academic publishers. 2003.

17. Papanastasiou, E. C.: A 'Rearrangement Procedure' for administering adaptive tests when review options are permitted. Doctoral dissertation, Michigan State University. 2001.

18. Laumer. S, von Stetten, A., Eckhardt, A.: E-Assessment / Bus Inf Syst Eng. 1 (3), 2009. P. 263–265. DOI: 10.1007/s12599-009-0051-6

19. Parshall, C. G., Spray, J. A., Kalohn, J. C., Davey, T.: Practical considerations in computer- based testing. NY : Springer. 2002.

20. Wainer, H.: CATs : Whither and whence. Psicologica, 21(1–2), 2000. P. 121–133.

21. Piton-Gonçalves J., Aluísio, S. M. An architecture for multidimensional computer adaptive test with educational purposes. ACM, New York, NY, USA, 2012. P. 17–24. DOI:10.1145/2382636.2382644

22. Piton-Gonçalves J., Aluísio, S.M.: Teste Adaptativo Computadorizado Multidimensional com propósitos educacionais: princípios e métodos in Ensaio Avaliação e Políticas Públicas em Educação 23 : P. 389–414, June, 2015. P. 59. DOI: 10.1590/S0104-40362015000100016

23. Kroker, L.: Vvedenie v klassicheskuyu i sovremennuyu teoriyu testov: uchebnik [Introduction to the Classical and Contemporary Theory of Tests: textbook]. M. : Logos, 2010. 668 p. (In Russian)

24. Avanesov, V.S.: Kompozitsiia testovykh zadanii [A Composition of Test Tasks] / V.S. Avanesov. M. : Tsentr testirovaniya, 2002. 239 p. (In Russian)

25. Vexler, V.A.: Pedagogicheskoe testirovanie dlya studentov, obuchayushchihsya po napravleniyu podgotovki 44.33.01 "Pedagogicheskoe obrazovanie" profil' "Informatika", ochnoi formy obucheniya: uchebno-metodicheskoe posobie [Pedagogical Testing for Full-Time Students of 44.33.01 Pedagogical Education Field of Study, Computer Science Specialty: study guide]. Saratov : SGU. 2015. 54 p. (In Russian)

26. Vexler, V.A., Otrokov, D.A.: Adaptivnoe testirovanie, kak vid obektivnogo kontrolya znanii, umenii i navykov obuchaemykh i odnogo iz sposobov povysheniya kachestva obrazovaniya [Adaptive Testing as a Kind of Objective Control of Learners' Knowledge and Skills and One Way of Improving the Quality of Education] // NOVAINFO.RU 2018, V. 1, no 94. P. 170–74. URL : Available at: http://novainfo.ru (In Russian)

27. Larin, S. N., Gerasimova, L. I., Gerasimov, E. V.: Adaptivnoe testirovanie urovnya znanii obuchaemykh kak instrumentarii realizatsii printsipov individualizatsii i differentsiatsii

obucheniya [Adaptive Testing of Learners' Knowledge Level as a Way of Implementing Individualization and Differentiation Approaches in Education] // Pedagogicheskii zhurnal. 2018, V. 8, no 2A. P. 48–57. (In Russian)

28. Zvonnikov, V.I.: Sovremennye sredstva otsenivaniya rezultatov obucheniya: uchebnik dlya stud. uchrezhdenii vyssh. prof. obrazovaniya [Contemporary Ways of Estimation of Academic Results: textbook for higher professional education schools] V. I. Zvonnikov M. B. Chelyshkova. M. : Izdatelskii tsentr "Akademiya", 2013. 304 p. (In Russian)

29. Efremova, N.F.: Testovyi kontrol' kachestva uchebnykh dostizhenii v obrazovanii [Test Check of the Quality of Academic Attainments in Education: textbook for higher professional education students]: Abstract of Doctor's degree dissertation] 13.00.01, Donskoi gosudarstvennyi pedagogicheskii universitet - Rostov on Don., 2003. – 25 p. (In Russian)

30. Majors, A. N.: Teoriya i praktika sozdaniya testov dlia sistemy obrazovaniya (Kak vybirat', sozdavat' i ispolzovat' testy dlia tselei obrazovaniya) [Theory and Practice of Tests Creation for Education System (How to select, to use and to create tests for academic purposes)]. M. : "Intellekt-tsentr", 2001. 296 p. (In Russian)

31. Minko, N.T.: Adaptivnoe testirovanie v usloviyakh personalnogo obrazovaniya [Adaptive Testing in Context of Personal Education]. Pedagogicheskie izmereniya. 2008. no 3. P. 95–102. (In Russian)

32. Koliada, M. G., Bugayova, T. I.: Vychislitel'naya pedagogika [Computational Pedagogy]. Rostov-on-Don, SFEDU, 2018. 270 p. (In Russian.)

33. Piager, J.: Structuralism. – New York : Basic Books, 1970. 453 p.

34. IMS Global Learning Consortium Announces the Next Generation for the Leading Digital Assessment Standard. [Electronic resource]. URL : http://www.imsglobal.org/article/ims-global-learning-consortium-announces-next-generation-leading-digital-assessment-standard

35. Krispin, L, Gregory, D.: Gibkoe testirovanie : prakticheskoe rukovodstvo dlia testirovshchikov PO i gibkikh komand [Flexible Testing: practical guidance for testers and flexible commands]. M. : OOO "I. D. Williams", 2010. 464 p. (In Russian.)

36. DeAyala, R. J., Koch, William R. A.: Computerized Implementation of a Flexilevel Test and Its Comparison with a Bayesian Computerized Adaptive Test, 1986. URL : https://files.eric.ed.gov/fulltext/ED269437.pdf

37. Luhn, Hans P.: Computer for Verifying Numbers, U.S. Patent 2 950 048, August 23. 1960.

38. Verhoeff, J.: Error detecting decimal codes (Ph. D.). Mathematical Centre Tracts 29, Amsterdam. 1969.

39. Damm, H. M.: Total anti-symmetrische Quasigruppen / Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.). Dem Fachbereich Mathematik und Informatik. Marburg/Lahn, 2004. 125 p.

40. Safonova, E. I.: Rekomendatsii po proektirovaniyu i ispolzovaniyu otsenochnykh sredstv pri realizatsii osnovnoi obrazovatelnoi programmy vysshego professionalnogo obrazovaniya (OOP VPO) novogo pokoleniya [Recommendations on Designing and Use of Assessment Means in View of the Basic Educational Program for Higher Professional Training of New Generation]. M. : RGGU, 2013. 75 p. (In Russian)