The Set Ordering Method for Scoring the Outcomes of Testing in Computerized Adaptive Testing

Simon Razmadze, Mariam Razmadze, Tornike Razmadze

Abstract: The given paper discusses an original method of the evaluation of outcomes of adaptive testing in the case of the strategy of the multilevel testing. The multiplicity/set of the outcomes of testing consists of atypical different-dimensional elements. The given paper defines the criteria of their comparison, describes the principles of ordering of the given multiplicity and draws the getting of a final score. The criterion of ordering of the outcomes of testing may not be the only option. The given paper illustrates this fact through a comparative discussion of two samples. An original procedure of testing is used for the presentation of the essence of the method. The given procedure is aimed to be illustrative, because a described method of assessment can be used for similar strategies. The next direction of the research implies the preparation of a new paper, which will show the usage of a method that is described in the given paper for the already-known approaches of a computerized adaptive testing.

Index Terms*:* adaptive testing with regression trees, applied mathematics, computerized adaptive testing, evaluation algorithm, multistage adaptive testing, ordering of a set, stradaptive testing.

1. A Discussed Model and Traditional Approaches

The described method in the given paper is based on the model where the multiplicity/set of the items of a test is divided into several parts. Each part is assigned a predefined level of difficulty. Subsequently, there is no other information available about the items on a test. In other words, the difficulty, discrimination and parameter of guessing for each item separately are not available.

The usage of the traditional CAT methods and standard algorithms in the conditions of such a little information is difficult. In order to obtain the advantages of the Item Response Theory (IRT), the tests should be designed, constructed, analyzed and interpreted within the framework of the given theory. Particularly, IRT implies that the ability of the particular examinee is known in advance. Based on these data, the parameters of the characteristic curve of items (difficulty, discrimination, guessing parameter) are determined [1].

The preliminary estimation of the abilities of an examinee and exquisite calibration of the set of the items is also the basis of non-IRT approaches to CAT [2].

Therefore, in contrast to the classical IRT concepts [3], [4], [1], Rasch's model [5], or non-IRT (i.e. the Measurement Decision Theory) of CAT [2], the model under discussion does not present the preliminary estimate parameter θ of an examinee's abilities and the items of the same level have the same difficulty. The final evaluation of an examinee is conducted using the scale 0-100.

Let us discuss some existing models that are based on splitting the items of a test into several parts and are close to the approach that is described in the given paper.

1.1 Stradaptive Testing

The stradaptive testing approach was used in the early period of the computerized adaptive testing. It was invented by Weiss [6], [7].

There are different strategies of the leveling, which were fundamentally discussed and studied earlier. These strategies are:

- Two-stage approach [8], [9], [10];
- Multi-Stage Approach:
  - Fixed Branching Models:
    - Pyramidal Strategy [10];
    - Flexilevel [11], [12], [13];
    - Stradaptive Testing [6], [7], [14];
  - Variable Branching Models:
    - Bayesian [7], [15];
    - Maximum likelihood approach [7].

The evaluation methods of a multi-level testing consider a number of correct answers, the difficulty of a final item, the difficulty of the $(n+1)^{th}$ item and an average difficulty of the items of a test. In case of the stradaptive testing approach [6] presents ten methods of the given type, while [7] adds 5 additional systems of evaluation. In all these systems the set of items is divided not only according to the levels, but each item within the level has its own parameters of difficulty and discrimination.

Table 1 presents the classical standard methods of scoring ([7], [14]) and explains why they can't be used for the model discussed in the given paper.

Table 1
The Methods of Estimation

N	The methods of scoring	Why not applicable
1	The difficulty of the most difficult item among the	According to the model discussed in this paper, there is no information about the difficulty of each item.
2	The difficulty of the (n+1)th item	
3	The difficulty of the most difficult non-chance answer	
4	The difficulty of the highest stratum of correct answers	
5	The difficulty of the (n+1)th stratum	According to the model discussed in this paper, the amount of levels of the test is not enough for the estimation of an examinee according to the scale 0-100.
6	The difficulty of the most difficult non-chance stratum	
7	The difficulty of an interpolated stratum	According to the model discussed in this paper, there is no information about the difficulty of each item.
8	The average difficulty of all correct items	
9	The average difficulty of correct items between ceiling and basal strata	
10	The average difficulty of correct items of the highest non-chance stratum	

1.2 Multistage Models

"Recently, multistage testing (MST) has been adopted by several important large-scale testing programs and become popular among practitioners and researchers" ([16], p. 104). "MST is a balanced compromise between linear test forms (i.e., paper-and-pencil testing and computer based testing) and traditional item-level computer-adaptive testing (CAT)" ([17], p. ii).

"In contrast to item-level CAT designs, which result in different test forms for each test taker, MST designs use a modularized configuration of preconstructed subtests and embedded score-routing schemes to prepackage validated test forms" ([18], p. 171).

The "stage" in multistage testing is an administrative division of the test that facilitates the adapting of the test to the examinee. Each examinee is administered modules for a minimum of two stages, where the exact number of stages is a test design decision affected by the extent of desired content coverage and measurement precision. In each stage, an examinee receives a module that is targeted in difficulty to the examinee's provisional ability estimate computed from performance on modules administered during the previous stage(s). Within a stage, there are typically two or more modules that vary from one another on the basis of average difficulty. Because the modules vary in this way, the particular sequence of item sets that any one examinee is presented with is adaptively chosen based on the examinee's ability estimate. After an examinee finishes each item set, his or her ability estimate is updated to reflect the new measurement information obtained about that examinee's ability, and the next module is chosen to provide an optimal level of measurement information for a person at that computed proficiency level. High-performing examinees receive modules of higher average difficulty, while less able examinees are presented with modules that are comparatively easier [19].

Thus, traditional CAT selects items for a test adaptively, while a multistage test provides a similar approach by using sets of items (modules, testlets) as "building blocks".

In order to build a panel using modules, a creator of a test uses linear programming or heuristic methods. Apart from this, the Maximum Fisher Information Method is used for obtaining the classification cut-points for the optimization of the information of a module [17]. All the above mentioned requires a specific knowledge. Our model does not have such limitations, because such a specific work is performed by a creator of an automatic system of testing, while a creator of a test has only to divide the testing items into several levels according to difficulty. This should not be difficult, because, we assume that a creator of a test is a professional in the field for which a test is created.

1.3 Regression Trees

Regression Trees and Tree Based Approaches use adaptive testing to predict the observed scores that test takers would have received if they had taken every item in a reference test or a pool [20].

"The tree-based MST algorithm provides efficient routing and scoring in constructing and analyzing MSTs without the aid of strong IRT Models, and thus it may be more robust to serious violations of the IRT model, allow item review within modules, and permit test developers to preconstruct modules for better content balance, while still providing efficient routing and accurate prediction" ([21], p. 171).

The calibration sample is needed in the case of adaptive testing IRT method, as well as in the case of using regression tree [20]. Calibration sample – is fixed amount of examinees, who have responded to all the questions of the test. For example, in order to obtain the calibration sample 250 examinees have responded to all the questions in the test [20]. In case of IRT method the calibration sample is used for the calibration of items, but in case of a regression tree – for building such tree, in other words, for determining the nodes of the tree.

Although there are no strong model assumptions, the regression tree approach has the important restriction that the sample used to calibrate the regression tree must be representative of the population for whom the test is designed. This restriction exists also for IRT-based CAT and MST when the item bank is built. Specifically, many items need to be calibrated on a representative testing sample ([21], p. 171).

The limitations described here require quite a lot of difficulties to be overcome by bringing the test to the condition of a practical use. In our model these problems do not exist.

2. The Importance of the Topic under Discussion

An accurately calibrated item bank is essential for a valid computerized adaptive test. However, in some cases, such as occupational testing, there is limited access to test takers for calibration. As a result of the limited access to possible test takers, collecting data to accurately calibrate an item bank in an occupational setting is usually difficult ([22], p. 2). In such situation, it might be reasonable to use a simplified method of testing, which is described in this paper. When can this simplified model be useful?

Let us imagine an automatic testing system that serves not only examinees, but the authors of the tests (experts). An expert who is registered in a system can create his/her own test, in other words, add his/her own items - questions and their suppositional answers - to the database. Later this test becomes available to other online users.

During the creation of an item pool an author is able to define the difficulty of each item himself/herself, in other words, the difficulty coefficient is estimated by experts. A computer system will specify these coefficients in future via the aggregated statistics. If we take into account the fact, that experts' qualification may be different (anyone can register in a system), an expert estimation of a test can be less reliable. Even for qualified experts the sorting of items of a test becomes more complicated when the number of items increases.

For this and other cases, when the calibration of the item pool is doubtful, it might be more reliable to split the pool of items into several levels of difficulty. This would be closer to reality and an expert's work would be eased. In fact, it is easier to split the item pool into several levels, e.g. into two (simple and difficult) or three (simple, medium, difficult) levels. The less is the amount of levels, the easier and more reliable is an expert estimation.

On the other hand, the more the levels are, the more informative is the item pool and the more informative the item pool is, the more perfect is the assessment of testing. This paper does not aim at defining the optimal number of levels according to different criteria. It provides the scoring method, which can correspond to different strategies in the case of a leveled pool of items.

In the framework of an automatic testing system, after working in a described mode and aggregating enough statistical data about the items of a test, an author is able to switch the test to IRT or other modern method in order to fully use the gathered information.

3. The Procedure of Testing

In the discussed model the set of items of a test is split into L parts. In order to simplify the explanation and due to the fact that the calculations are made for the item pool split into 5 levels, without touching the generalization, L is assumed to be equal to 5.

Let us denote the levels of the item pool from 1 to 5 in accordance with the increase of difficulty. During the testing the first item is selected from the items of the third level by means of a random number generator. In case of a correct answer to the first question, the second question is selected from the fourth level. In case of an incorrect answer, the second question is selected from the second level.

Generally, in case of a correct answer, the next item is selected from one step higher level and in case of an incorrect answer – from one step lower level. If there is no higher (lower) level left the next item is selected from the same level.

The last item is given, when a person gets to the same level for the fifth time. The examinee gets the fifth item from this level and after answering it the testing process is over.

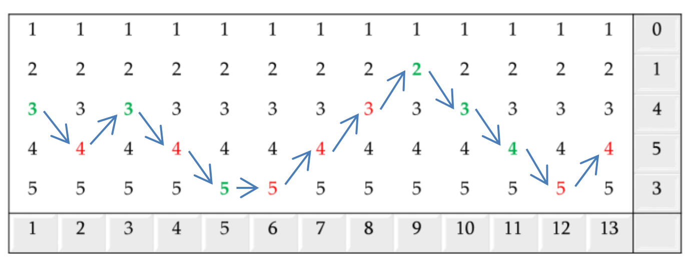

Fig. 1. The sample of the outcomes of testing.

Figure 1 describes the sample of an outcome of testing. Horizontally the question number is measured, while vertically a level of the difficulty of an item is presented. The data of the last column shows how many questions were provided per level. Testing was stopped after the 13th question, because at this moment the examinee had reached the 4th level for the 5th time.

Theoretically, the number of questions asked during a testing session can range from seven (none of the answers were correct) to twenty one (on each level four questions were asked and the 5th additional item was presented from a corresponding level). Practically, the test should end after 9-10 questions.

4. The Systems of Evaluation

The given paper presents 100-point system of the outcomes of testing in accordance to the difficulty of questions and correctness of their answers.

After defining the score within 100-point system, the realization of any - less than 100-point - system (five-point, ten-point, etc.) becomes easier. For instance, in case of the evaluation within the system with the grades A, B, C, D, E, FX, F:

- 91-100 – A;
- 81-90 - B;
- 71-80 - C;
- 61-70 - D;
- 51–60 - E;
- 41-50 - FX – did not pass, but an examinee is given a possibility to go across the final exam;
- 0-40 – F – failed, an examinee must attend the course again in order to obtain a credit.

5. Description of the Outcomes of Testing

In order to evaluate the outcomes of testing (for writing any score) each outcome must be marked by an integer from the range 0-100. The higher number must correspond to the better outcome of testing.

Prior to making the correspondence with numbers, it is important to define the way of the comparison of the outcomes of testing and the way of the determination of the "better" and "worse" ones. Sometimes it seems clear from the first sight, but in other occasions it turns out to be quite difficult. For example, which outcome of testing is better – the one with 5 incorrect and 10 correct answers or the one with 6 incorrect and 11 correct answers? Especially, if we take into account different levels of difficulty for items, the comparison of the outcomes of testing will not be easy.

5.1 The Weight of an Item

It's obvious, that the outcomes of testing should be evaluated according to the difficulty of the given items and the correctness of the answers. First of all, different coefficients (in other words – "weight") should be assigned to each item with different difficulty. The weight for an item can be defined according to its level of difficulty. Accordingly, the first level items will be "weight" 1, the second level – 2, the third level – 3, the fourth level – 4 and the fifth level – 5.

5.2 General View of the Outcomes of Testing

The outcome of testing is the vector of numbers that correspond to the levels of the items given during the testing process. This vector consists of two parts – the sets of incorrect and correct answers.

For example, if an examinee could not provide the correct answer to 3 questions and these items were from levels five, three and two, then the set of incorrect answers would be {5,3,2}. If the examinee provided correct answers to 9 questions on the same test, we would get the multiplicity of correct answers via considering the levels of the correct answers, for instance, {5,5,5,5,4,4,3,2,1}.

The unity of the multiplicity of incorrect and correct answers (in the above described example {5,3,2} and {5,5,5,5,4,4,3,2,1}) form one whole object called the outcome of testing. Let us denote this unity as: {5,3,2 & 5,5,5,5,4,4,3,2,1}.

Generally, the outcome of testing is element of $\{m_1, m_2, \dots, m_\mu \,\&\, r_1, r_2, \dots, r_\rho\}$ type, where μ is the amount of incorrect answers, m_i, $i = \overline{1,\mu}$ - the level of an i^{th} incorrect answer, ρ – the number of correct answers, and $r_j, j = \overline{1,\rho}$ – the level of a j^{th} correct answer.

Let us denote a set of outcomes of testing by N. Table 2 consists of the samples of real outcomes of testing. Due to the fact that the number of items given during one session can vary from 7 to 21, the set N consists of 7-21 different-dimensional vectors within the framework of the discussed procedure of testing.

Table 2
The samples of the outcomes of testing

Question Levels	Incorrect Answers	Correct Answers	Testing Outcom
3455555		5,5,5,5,5,4,3	& 5,5,5,5,5,4,3
34555545	5	5,5,5,5,4,4,3	5 & 5,5,5,5,4,4,3
3455555	5	5,5,5,5,4,3	5 & 5,5,5,5,4,3
343455555	4	5,5,5,5,5,4,3,3	4 & 5,5,5,5,5,4,3,3
345554545	5,5	5,5,5,4,4,4,3	5,5 & 5,5,5,4,4,4,3
34555545	5,5	5,5,5,4,4,3	5,5 & 5,5,5,4,4,3
3455554345	5,4	5,5,5,5,4,4,3,3	5,4 & 5,5,5,5,4,4,3,3
323455555	3	5,5,5,5,5,4,3,2	3 & 5,5,5,5,5,4,3,2
34343455555	4,4	5,5,5,5,5,4,3,3,3	4,4 & 5,5,5,5,5,4,3,3,3
343455555	5,4	5,5,5,5,4,3,3	5,4 & 5,5,5,5,4,3,3
3455454545	5,5,5	5,5,4,4,4,4,3	5,5,5 & 5,5,4,4,4,4,3
34555454345	5,5,4	5,5,5,4,4,4,3,3	5,5,4 & 5,5,5,4,4,4,3,3

We created the computer program that gives the complete set N of outcomes of testing. In case of the five-level items and above described procedure of testing the set N consists of 1432 elements.

6. The Criteria of Evaluation of the Outcomes of Testing

6.1 Stages of Evaluation of the Outcomes of Testing

Our final aim is the assessment of the outcome of testing, which can be achieved through the following two stages:

1. Ordering of the set of the outcomes of testing. This stage assumes the construction of ideas about the better/worse outcomes of testing as well as the reflection of the set of the outcomes of testing on the number set in order to form an increasing function (in other words, the higher number corresponds to the better outcomes of testing);
2. The transformation of the derived number set matching the outcomes of testing into the segment of integers [0,100]. In other words, assigning 100-point scoring system to the elements of an ordered set.

6.2 Parameters of the Evaluation of the Outcomes of Testing

The outcomes of testing can be evaluated according to different opinions. Herewith, the evaluations performed with different logic can be different. This fact does not exclude that each opinion can have its own objective validity. Afterwards, we will discuss the set of the outcomes of testing based on two different criteria, which ensures a fair evaluation of the better/worse outcomes of testing. The following parameters should be taken into account for the realization of the criteria of assessment:

- a weighted sum of points of correct answers;
- a weighted sum of points of incorrect answers;
- an average difficulty of incorrect answers.

A weighted sum of points of correct answers considers the number of correct answers, while the weight of incorrect answers implies the number of errors, which are also important for the evaluation of the outcomes of testing. It is also possible to operate with the average meanings of correct and incorrect answers. However, these average meanings can be calculated by dividing the weighted sum of points into the number of corresponding answers.

Let us create the formulas for the calculation of a weighted sum of points of correct and incorrect answers and an average difficulty of incorrect answers, that will be used in the criteria of an estimation of the outcomes of testing.

6.2.1 A weighted sum of the points of correct answers.

First of all, a weighted sum of points of correct answers should be taken into account for the evaluation of the outcomes of testing. Let's denote it with the letter R.

A weighted sum of points of correct answers R of the outcomes of testing $n=\{m_1,m_2,\dots,m_\mu \,\&\, r_1,r_2,\dots,r_\rho\}$ can be calculated via using the following formula:

$$R(n)=\sum_{l=1}^{5}(l*\mathrm{c}_l)=\sum_{j=1}^{\rho}\mathrm{r}_j\,,\ \ n\in N \qquad (1),$$

where l is the level of the item and c_l is a number of correct answers to the questions of the level l. This is similar to give an examinee the point corresponding to the level of difficulty of the item and then summing up the results.

6.2.2 A weighted sum of points of incorrect answers.

The difference between calculating a weighted sum of points of incorrect and correct answers lies in the fact that in case of incorrect answers the weight of the item in the weighted sum should be considered as inversely proportional to the level of the item:

- For the items of the 1st level, the weight of an incorrect item would be 5;
- For the items of the 2nd level – 4;
- For the items of the 3rd level – 3;
- For the items of the 4th level – 2;
- For the items of the 5th level – 1.

The greater is a weighted sum of the levels of mistakes, the less is the score of the evaluation, despite the number of incorrect answers.

A weighted sum of points of incorrect answers R of the outcomes of testing $n=\{m_1,m_2,\dots,m_\mu \,\&\, r_1,r_2,\dots,r_\rho\}$ can be calculated via using the following formula:

$$M(n)=\sum_{\mathrm{l}=1}^{5}(6-\mathrm{l})*\mathrm{d}_\mathrm{l}=\sum_{j=1}^{\mu}(6-\mathrm{m}_j)\,,\ \ n\in N \qquad (2),$$

where l is the level of the item and d_l is a number of incorrect answers to the questions of the level l. This is similar to give an examinee the point corresponding to the level of difficulty of the item and then summing up the results.

6.2.3 An average difficulty of incorrect answers.

The estimation of the outcomes of testing requires the analysis of errors too. The more is an average level of incorrect answers, the more difficult test was passed by an examinee.

An average difficulty of incorrect answers A of the outcomes of testing $n=\{m_1,m_2,\dots,m_\mu \,\&\, r_1,r_2,\dots,r_\rho\}$ can be calculated via the following formula:

$$A(n)=\frac{1}{\mu}\sum_{i=1}^{\mu}m_i\,,\ \ n\in N \qquad (3).$$

6.3 The First Criterion of the Evaluation of the Outcomes of Testing

Let us state the first criterion for the evaluation of the outcomes of testing:

- The greater is the point of incorrect answers M(n), the worse are the outcomes of testing (the less score should be written). As a result of this consideration, the difficulty and the number of items corresponding to incorrect answers will be taken into account;
- The greater is the weighted sum of correct answers R(n), the better are the outcomes of testing (the higher score should be written). As a result, the difficulty and the number of items corresponding to correct answers will be taken into account.

According to the given considerations, the score of the outcomes of testing is directly proportional to a weighted sum of points of correct answers and is inversely proportional to a weighted sum of points of incorrect answers.

According to the first criterion of the evaluation, the following formula is used for the calculation of the score S of the outcomes of testing $n = \{m_1, m_2, \dots, m_\mu \& r_1, r_2, \dots, r_\rho\}$:

$$S(n) = \frac{R}{1+M}, \quad n \in N \qquad (4),$$

The purpose of adding 1 in the denominator aims at avoiding the division on 0.

The following can be obtained by taking into account the formulas (1) and (2):

$$S(n) = \frac{\sum_{l=1}^{5}(l * c_l)}{1 + \sum_{l=1}^{5}(6-l) * d_l}, \quad n \in N \qquad (5), \text{ or}$$

$$S(n) = \frac{\sum_{l=1}^{5}(l * c_l)}{1 + \sum_{j=1}^{\mu}(6 - m_j)}, \quad n \in N \qquad (6), \text{ or}$$

$$S(n) = \frac{\sum_{j=1}^{\rho} r_j}{1 + \sum_{l=1}^{5}(6-l) * d_l}, \quad n \in N \qquad (7), \text{ or}$$

$$S(n) = \frac{\sum_{j=1}^{\rho} r_j}{1 + \sum_{j=1}^{\mu}(6 - m_j)}, \quad n \in N \qquad (8).$$

6.4 The Second Criterion of the Evaluation of the Outcomes of Testing

The second criterion for the evaluation of the outcomes of testing considers the following:

- The greater is a weighted sum of correct answers R(n), the better are the outcomes of testing (the higher score should be written). As a result of this consideration the difficulty and the number of items corresponding to the correct answers will be taken into account;
- The items corresponding to incorrect answers must influence on a final result. Therefore:
  - The greater is an average difficulty of incorrect answers A, the higher is the final score;
  - An average difficulty A(n) does not take into account the number of incorrect answers. For example, there may be one error on the 3rd level (average - 3), or three errors on the 2nd, 3rd and 4th levels (average - 3). These two cases are not equivalent, apparently, the higher score must be written for one error. Therefore, the less is the number of incorrect answers μ, the higher is a final score.

According to the second criterion of the evaluation, the following formula is used for the calculation of the score F of the outcomes of testing $n = \{m_1, m_2, \dots, m_\mu \; \& \; r_1, r_2, \dots, r_\rho\}$:

$$F(n) = R * \frac{A}{\mu}, \quad n \in N \qquad (9).$$

The following can be obtained by taking into account the formulas (1) and (3):

$$F(n) = \frac{1}{\mu^2} * \sum_{i=1}^{\mu} m_i * \sum_{l=1}^{5}(l * c_l), \quad n \in N \qquad (10), \text{ or}$$

$$F(n) = \frac{1}{\mu^2} * \sum_{i=1}^{\mu} m_i * \sum_{j=1}^{\rho} r_j, \quad n \in N \qquad (11).$$

7. Ordering of the Set of the Outcomes of Testing

7.1 Ordering of the Outcomes of Testing According to the First Criterion

According to the formulas (4)-(8), each element of the set of the outcomes of testing N corresponds to a particular number. The rule of ordering of the set N:

For two elements n_1 and n_2 from the set N, it can be said that $n_1 \le n_2$, if $S(n_1) \le S(n_2)$.

The rule of ordering of a set must satisfy the properties of reflexivity, transitivity and antisymmetry [23], [24].

1. Reflexivity:
   $For\ \forall n \in N\ there\ is\ S(n) \le S(n). Therefore, n \le n$.
2. Transitivity:
   $if\ n_1 \le n_2$ and $n_2 \le n_3$, then $S(n_1) \le S(n_2)$ and $S(n_2) \le S(n_3)$, so it comes out that $S(n_1) \le S(n_3)$, this means $n_1 \le n_3$.
3. Anti-symmetry. There must be:
   $if\ n_1 \le n_2$ and $n_2 \le n_1$, $\Rightarrow n_1 = n_2$.

This condition is not met, because S(n) is only a reflection and not one-to-one correspondence. Therefore, two different outcomes of testing may have the same reflection. For example:

$$n_1 = \{4,3,2 \;\&\; 5,5,5,5,5,4,3,3,2,1\}$$
$$n_2 = \{3,3,3 \;\&\; 5,5,5,5,5,4,3,2,2,2\}$$

According to the formula (5):

$$S(n_1) = S(\{4,3,2 \;\&\; 5,5,5,5,5,4,3,3,2,1\}) = \frac{1*1+2*1+3*2+4*1+5*5}{1+(6-2)*1+(6-3)*1+(6-4)*1} = \frac{38}{10}.$$

$$S(n_2) = S(\{3,3,3 \;\&\; 5,5,5,5,5,4,3,2,2,2\}) = \frac{2*3+3*1+4*1+5*5}{1+(6-3)*3} = \frac{38}{10}.$$

Despite the fact that the condition $n_1 \le n_2$ and $n_2 \le n_1$ is met, the outcomes of testing n_1 *and* n_2 are different. This contradicts to the anti-symmetry.

The problem will be solved if such elements are considered as equal. This fact will not distract the solution of our main task, because there are no restrictions for evaluating two different (but similar) outcomes of testing with the same score.

Therefore, let us consider that the outcomes of testing that have the same reflection are equal. As a result of this consideration, the condition of the anti-symmetry will be met and the set N will be totally ordered.

Table 3
The First 20 and Last 20 Points of the Outcomes of Testing According to the First Criterion

N	Question Levels	Incorrect Answers	Correct Answers	Score
1	3455555		5,5,5,5,5,4,3	32.00
2	34555545	5	5,5,5,5,4,4,3	15.50
3	3455555	5	5,5,5,5,4,3	13.50
4	343455555	4	5,5,5,5,5,4,3,3	11.67
5	345554545	5,5	5,5,5,4,4,4,3	10.00
6	34555545	5,5	5,5,5,4,4,3	8.67
7	3455554345	5,4	5,5,5,5,4,4,3,3	8.50
8	323455555	3	5,5,5,5,5,4,3,2	8.50
9	34343455555	4,4	5,5,5,5,5,4,3,3,3	7.60
10	343455555	5,4	5,5,5,5,4,3,3	7.50
11	3455454545	5,5,5	5,5,4,4,4,4,3	7.25
12	34555454345	5,5,4	5,5,5,4,4,4,3,3	6.60
13	3234555545	5,3	5,5,5,5,4,4,3,2	6.60
14	345554545	5,5,5	5,5,4,4,4,3	6.25
15	345555434345	5,4,4	5,5,5,5,4,4,3,3,3	6.17
16	34323455555	4,3	5,5,5,5,5,4,3,3,2	6.17
17	343434345555	4,4,4	5,5,5,5,5,4,3,3,3,3	5.86
18	3455554345	5,5,4	5,5,5,4,4,3,3	5.80
19	323455555	5,3	5,5,5,5,4,3,2	5.80
20	34343455555	5,4,4	5,5,5,5,4,3,3,3	5.50

N	Question Levels	Incorrect Answers	Correct Answers	Score
1413	32323211111	3,3,3,2,1,1,1,1	2,2,1	0.15
1414	323232321111 1	3,3,3,3,2,1,1,1,1,1	2,2,2	0.14
1415	343211111	4,3,2,1,1,1,1	3,1	0.13
1416	3232121111	3,3,2,2,1,1,1	2,1,1	0.13
1417	32323212111 1	3,3,3,2,2,1,1,1,1	2,2,1	0.13
1418	34323211111	4,3,3,2,1,1,1,1,1	3,2	0.13
1419	3432121111	4,3,2,2,1,1,1,1	3,1	0.12
1420	32321212111	3,3,2,2,2,1,1,1	2,1,1	0.12
1421	321212111	3,2,2,2,1,1	1,1,1	0.12
1422	32323211111	3,3,3,2,1,1,1,1,1	2,2	0.10
1423	3212121211	3,2,2,2,2,1,1	1,1,1	0.10
1424	323211111	3,3,2,1,1,1,1	2,1	0.10
1425	3232121111	3,3,2,2,1,1,1,1	2,1	0.09
1426	343211111	4,3,2,1,1,1,1,1	3	0.09
1427	32121111	3,2,2,1,1,1	1,1	0.07
1428	321212111	3,2,2,2,1,1,1	1,1	0.06
1429	323211111	3,3,2,1,1,1,1,1	2	0.06
1430	3211111	3,2,1,1,1,1	1	0.04
1431	32121111	3,2,2,1,1,1,1	1	0.03
1432	3211111	3,2,1,1,1,1,1		0.00

Table 3 shows the first 20 and last 20 scores of the outcomes of testing obtained after ordering in accordance with the first criterion.

7.2 The Final Score of the Outcomes of Testing According to the First Criterion

The first stage of the evaluation of the outcomes of testing is passed. The main problem – ordering of the set N - is solved. The elements of the set N are ordered according to the increasing reflections of the function S. If an increasing condition is not abolished, the meanings of these reflections can be changed arbitrarily. However, the ordering of the set N will not change.

Let us move to the second stage. It considers the conversion of the number set obtained by the outcomes of testing into the segment [0,100] of integer numbers, in other words, it considers assigning the points to the elements of the obtained set in accordance with the 100-point system. For reaching the above mentioned, it is needed to correct the data in the column "Score" (without changing the order) and select the incremental function, which will transform the reflection of the function S into the segment [0,100]. An incremental property of a function ensures holding the principle "the better the outcomes of testing – the higher the score".

If we observe the points of the outcomes of testing (the reflections of the function S), we will see that their meanings vary from min=0 to max=32. The most important is the fact that the meanings of the points in the area of maximum significantly differ from each other in contrast to the points in any other area. Accordingly, during the transformation through a linear or a non-linear analytical function the high scores are given rarely or are not given at all.

The first 90 points were artificially changed by equally decreasing numbers, step 0.03. This fact significantly improved a non-equivalence of the data - min=0 and max=5.46 were obtained. The ordering of the outcomes of testing was not changed.

After the correction the reflections of the function S were transformed into the segment [0,100] through a logarithmic function:

$$S_{(100)} = int[100 * log_2(S_{(5.46)} + 1 - \min)/(log_2(max + 1 - \min)]$$

Table 4
The First 20 and Last 20 Corrected Points of the Outcomes of Testing According to the First Criterion

N	Question Levels	Incorrect Answers	Correct Answers	Final Score
1	3455555		5,5,5,5,5,4,3	100
2	34555545	5	5,5,5,5,4,4,3	99
3	3455555	5	5,5,5,5,4,3	99
4	343455555	4	5,5,5,5,5,4,3,3	99
5	345554545	5,5	5,5,5,4,4,4,3	98
6	34555545	5,5	5,5,5,4,4,3	98
7	3455554345	5,4	5,5,5,5,4,4,3,3	98
8	323455555	3	5,5,5,5,5,4,3,2	98
9	34343455555	4,4	5,5,5,5,5,4,3,3,3	97
10	343455555	5,4	5,5,5,5,4,3,3	97
11	3455454545	5,5,5	5,5,4,4,4,4,3	97
12	34555454345	5,5,4	5,5,5,4,4,4,3,3	97
13	3234555545	5,3	5,5,5,5,4,4,3,2	96
14	345554545	5,5,5	5,5,4,4,4,3	96
15	345555434345	5,4,4	5,5,5,5,4,4,3,3,3	96
16	34323455555	4,3	5,5,5,5,5,4,3,3,2	96
17	343434345555	4,4,4	5,5,5,5,5,4,3,3,3,3	95
18	3455554345	5,5,4	5,5,5,4,4,3,3	95
19	323455555	5,3	5,5,5,5,4,3,2	95
20	34343455555	5,4,4	5,5,5,5,4,3,3,3	95

N	Question Levels	Incorrect Answers	Correct Answers	Final Score
1413	32323211111	3,3,3,2,1,1,1,1	2,2,1	7
1414	3232323211111	3,3,3,3,2,1,1,1,1,1	2,2,2	7
1415	343211111	4,3,2,1,1,1,1	3,1	6
1416	3232121111	3,3,2,2,1,1,1	2,1,1	6
1417	323232121111	3,3,3,2,2,1,1,1,1	2,2,1	6
1418	34323211111	4,3,3,2,1,1,1,1,1	3,2	6
1419	3432121111	4,3,2,2,1,1,1,1	3,1	5
1420	32321212111	3,3,2,2,2,1,1,1	2,1,1	5
1421	321212111	3,2,2,2,1,1	1,1,1	5
1422	32323211111	3,3,3,2,1,1,1,1,1	2,2	5
1423	3212121211	3,2,2,2,2,1,1	1,1,1	5
1424	323211111	3,3,2,1,1,1,1	2,1	4
1425	3232121111	3,3,2,2,1,1,1,1	2,1	4
1426	343211111	4,3,2,1,1,1,1,1	3	4
1427	32121111	3,2,2,1,1,1	1,1	3
1428	321212111	3,2,2,2,1,1,1	1,1	3
1429	323211111	3,3,2,1,1,1,1,1	2	2
1430	3211111	3,2,1,1,1,1	1	1
1431	32121111	3,2,2,1,1,1,1	1	1
1432	3211111	3,2,1,1,1,1,1		0

Table 4 shows the first 20 and last 20 corrected points of the outcomes of testing according to the first criterion. Visually, the obtained results are shown in the graph, which is presented in Figure 2.

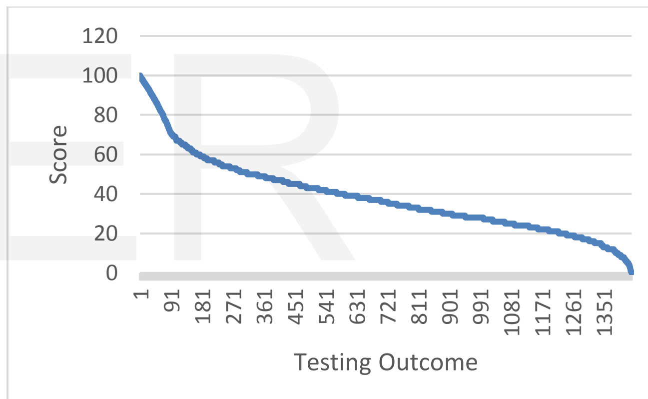

Fig. 2. The graph of the outcomes of testing ordered according to the first criterion.

7.3 Ordering of the Outcomes of Testing According to the Second Criterion

According to the formulas (9)-(11), each element of the set N of the outcomes of testing corresponds to a particular number. The rule of ordering of the set N:

For two elements n_1 and n_2 from the set N, it can be said that $n_1 \leq n_2$, if $F(n_1) \leq F(n_2)$.

Despite the fact, that the following discussion is analogous to the one held during the ordering according to the first criterion, that has to be held, because we have to deal with a different function. Table 5 shows the first 20 and last 20 points of the outcomes of testing after ordering according to the second criterion.

International Journal of Scientific & Engineering Research Volume 8, Issue 5, May-2017
ISSN 2229-5518

Table 5
The First 20 and Last 20 Points of the Outcomes of Testing According to the Second Criterion

N	Question Levels	Incorrect Answers	Correct Answers	Score
1	3455555		5,5,5,5,5,4,3	160.00
2	34555545	5	5,5,5,5,4,4,3	155.00
3	343455555	4	5,5,5,5,5,4,3,3	140.00
4	3455555	5	5,5,5,5,4,3	135.00
5	323455555	3	5,5,5,5,5,4,3,2	102.00
6	3455554345	5,4	5,5,5,5,4,4,3,3	76.50
7	34343455555	4,4	5,5,5,5,5,4,3,3,3	76.00
8	345554545	5,5	5,5,5,4,4,4,3	75.00
9	343455555	5,4	5,5,5,5,4,3,3	67.50
10	3234555545	5,3	5,5,5,5,4,4,3,2	66.00
11	34555545	5,5	5,5,5,4,4,3	65.00
12	34323455555	4,3	5,5,5,5,5,4,3,3,2	64.75
13	323455555	5,3	5,5,5,5,4,3,2	58.00
14	3434343455555	4,4,4	5,5,5,5,5,4,3,3,3,3	54.67
15	32323455555	3,3	5,5,5,5,5,4,3,2,2	54.00
16	345555434345	5,4,4	5,5,5,5,4,4,3,3,3	53.44
17	34555454345	5,5,4	5,5,5,4,4,4,3,3	51.33
18	3434323455555	4,4,3	5,5,5,5,5,4,3,3,3,2	48.89
19	3455454545	5,5,5	5,5,4,4,4,4,3	48.33
20	345555432345	5,4,3	5,5,5,5,4,4,3,3,2	48.00

N	Question Levels	Incorrect Answers	Correct Answers	Score
1413	32323211111	3,3,3,2,1,1,1,1	2,2,1	1.17
1414	3232323211111	3,3,3,3,2,1,1,1,1,1	2,2,2	1.14
1415	343211111	4,3,2,1,1,1,1	3,1	1.06
1416	3232121111	3,3,2,2,1,1,1	2,1,1	1.06
1417	323232121111	3,3,3,2,2,1,1,1,1	2,2,1	1.05
1418	34323211111	4,3,3,2,1,1,1,1,1	3,2	1.05
1419	3432121111	4,3,2,2,1,1,1,1	3,1	0.94
1420	32321212111	3,3,2,2,2,1,1,1	2,1,1	0.94
1421	321212111	3,2,2,2,1,1	1,1,1	0.92
1422	3212121211	3,2,2,2,2,1,1	1,1,1	0.80
1423	32323211111	3,3,3,2,1,1,1,1,1	2,2	0.79
1424	323211111	3,3,2,1,1,1,1	2,1	0.73
1425	3232121111	3,3,2,2,1,1,1,1	2,1	0.66
1426	343211111	4,3,2,1,1,1,1,1	3	0.66
1427	32121111	3,2,2,1,1,1	1,1	0.56
1428	321212111	3,2,2,2,1,1,1	1,1	0.49
1429	323211111	3,3,2,1,1,1,1,1	2	0.41
1430	3211111	3,2,1,1,1,1	1	0.25
1431	32121111	3,2,2,1,1,1,1	1	0.22
1432	3211111	3,2,1,1,1,1,1		0.00

7.4 The Final Score of the Outcomes of Testing According to the Second Criterion

The first stage of the evaluation of the outcomes of testing is passed. The main problem – ordering of the set N - is solved. The elements of the set N are ordered according to the increasing reflections of the function F. If an increasing condition is not abolished, the meanings of these reflections can be changed arbitrarily. However, the ordering of the set N will not change.

Let us move to the second stage. It considers the conversion of the number set obtained by the outcomes of testing into the segment [0,100] of integer numbers, in other words, it considers assigning the points to the elements of the obtained set in accordance with the 100-point system. For reaching the above mentioned, it is needed to correct the data in the column "Score" (without changing the order) and select the incremental function, which will transform the reflection of the function F into the segment [0,100]. An incremental property of a function ensures holding the principle "the better the outcomes of testing – the higher the score".

If we observe the points of the outcomes of testing (the reflections of the function F), we will see that their meanings vary from min=0 to max=160. The most important is the fact that the meanings of the points in the area of maximum significantly differ from each other in contrast to the points in any other area. Accordingly, during the transformation using a linear or a non-linear analytical function the high scores are given rarely or are not given at all.

The first 30 points were artificially changed by equally decreasing numbers, step 0.5. This fact significantly improved a non-equivalence of the data - min=0 and max=53 were obtained. The ordering of the outcomes of testing was not changed.

After the correction the reflections of the function F were transformed into the segment [0,100] through a logarithmic function:

$$F_{(100)} = int[100 * log_2(F_{(53)} + 1 - \min)/(log_2(max + 1 - \min)]$$

Table 6
The First 20 and Last 20 Corrected Points of the Outcomes of Testing According to the Second Criterion

N	Question Levels	Incorrect Answers	Correct Answers	Final Score
1	3455555		5,5,5,5,5,4,3	100
2	34555545	5	5,5,5,5,4,4,3	99
3	343455555	4	5,5,5,5,5,4,3,3	99
4	3455555	5	5,5,5,5,4,3	99
5	323455555	3	5,5,5,5,5,4,3,2	99
6	3455554345	5,4	5,5,5,5,4,4,3,3	98
7	34343455555	4,4	5,5,5,5,5,4,3,3,3	98
8	345554545	5,5	5,5,5,4,4,4,3	98
9	343455555	5,4	5,5,5,5,4,3,3	98
10	3234555545	5,3	5,5,5,5,4,4,3,2	97
11	34555545	5,5	5,5,5,4,4,3	97
12	34323455555	4,3	5,5,5,5,5,4,3,3,2	97
13	323455555	5,3	5,5,5,5,4,3,2	97
14	3434343455555	4,4,4	5,5,5,5,5,4,3,3,3,3	96
15	32323455555	3,3	5,5,5,5,5,4,3,2,2	96
16	345555434345	5,4,4	5,5,5,5,4,4,3,3,3	96
17	34555454345	5,5,4	5,5,5,4,4,4,3,3	95
18	3434323455555	4,4,3	5,5,5,5,5,4,3,3,3,2	95
19	3455454545	5,5,5	5,5,4,4,4,4,3	95
20	345555432345	5,4,3	5,5,5,5,4,4,3,3,2	95

N	Question Levels	Incorrect Answers	Correct Answers	Final Score
1413	32323211111	3,3,3,2,1,1,1,1	2,2,1	16
1414	3232323211111	3,3,3,3,2,1,1,1,1,1	2,2,2	16
1415	343211111	4,3,2,1,1,1,1	3,1	15
1416	3232121111	3,3,2,2,1,1,1	2,1,1	14
1417	323232121111	3,3,3,2,2,1,1,1,1	2,2,1	14
1418	34323211111	4,3,3,2,1,1,1,1,1	3,2	13
1419	3432121111	4,3,2,2,1,1,1,1	3,1	12
1420	32321212111	3,3,2,2,2,1,1,1	2,1,1	11
1421	321212111	3,2,2,2,1,1	1,1,1	10
1422	3212121211	3,2,2,2,2,1,1	1,1,1	10
1423	32323211111	3,3,3,2,1,1,1,1,1	2,2	9
1424	323211111	3,3,2,1,1,1,1	2,1	8
1425	3232121111	3,3,2,2,1,1,1,1	2,1	6
1426	343211111	4,3,2,1,1,1,1,1	3	7
1427	32121111	3,2,2,1,1,1	1,1	5
1428	321212111	3,2,2,2,1,1,1	1,1	4
1429	323211111	3,3,2,1,1,1,1,1	2	3
1430	3211111	3,2,1,1,1,1	1	2
1431	32121111	3,2,2,1,1,1,1	1	1
1432	3211111	3,2,1,1,1,1,1		0

Table 6 shows the first 20 and last 20 corrected points of the outcomes of testing according to the second criterion. Visually, the obtained results are shown in the graph, which is presented in Figure 3.

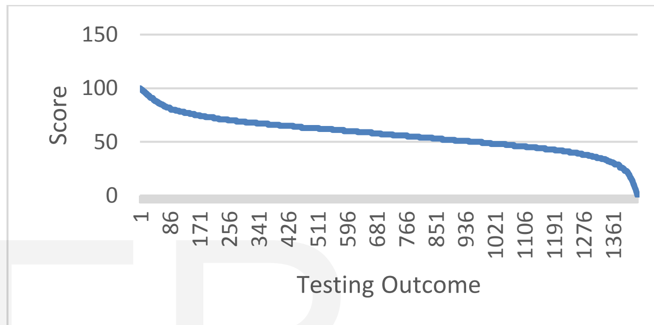

Fig. 3. The graph of estimation of the outcomes of testing ordered according to the second criterion.

7.5 Comparative Analysis of Ordering of Results

After the comparison of the tables 4 and 6 it became obvious that the results obtained according to the first and second criteria differ from each other. For example, according to the first criterion the outcomes of testing $n_1 = \{3\ \&\ 5,5,5,5,5,4,3,2\}$ and $n_2 = \{5,5\ \&5,5,5,4,4,4,3\}$ are estimated with the same point (98), while according to the second criterion $n_1 = \{3\ \&\ 5,5,5,5,5,4,3,2\}$ (99) is better than $n_2 = \{5,5\ \&5,5,5,4,4,4,3\}$ (98).

Simply, the question is as follows: which is more important - one error of the middle (third) level or two errors of the difficult (fifth) level? The given question and generally, the analysis of the obtained tables are the subject of another research. A creator of an automatized system of testing can make his/her decision regarding the above mentioned.

Moreover, besides the discussed criteria one can create a lot of other criteria. Each of them will have its own table of results. In all cases the issue should be solved according to the considerations of the creator of an automatized system of testing – he/she has to decide the logic of which criterion is more relevant and which final result is acceptable.

8. Conclusions

The set ordering method for scoring the outcomes of testing can be used in cases of different procedures. Creator of a test has no direct contact with this method and its specific nuances, for the realization of method is an one-time procedure in the computerized adaptive testing system.

The set ordering method for scoring the outcomes of testing is oriented on a creator of a test. It simplifies a creator's workflow. The method does not require a detailed calibration of an item pool or a preliminary testing for the creation of a calibration sample. A preliminary workflow of a creator of a test may consist of only splitting items into several levels of difficulty on the basis of an expert estimation.

In the conditions of the lack of the information about the items of a test and level of examinees' knowledge, the given method maximally uses an existed information for evaluating an examinee: every answer to the item provided by an examinee is taken into account, a set of responses is compared with all possible variations and is placed on a corresponding level of a hierarchy of scoring.

The next direction of the research implies the preparation of a new paper, which will show the usage of a method that is described in the given paper for the already-known approaches of a computerized adaptive testing.

9. References

[1] F. Baker, "The Basics of Item Response Theory". *ERIC Clearinghouse on Assessment and Evaluation*, Maryland, MD: University of Maryland, College Park, 2001. Retrieved from http://echo.edres.org:8080/irt/baker;

[2] L.M. Rudner, "Scoring and Classifying Examinees using Measurement Decision Theory". *Practical Assessment, Research & Evaluation*. Volume 14, Number 8, 2009. Retrieved from http://echo.edres.org:8080/mdt;

[3] S.E. Embretson & S.P. Reise, "Item Response Theory for Psychologists". *Multivariate Applications Books Series.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers, 2000;

[4] W.J. Van der Linden & R.K. Hambleton, "Handbook of Item Response Theory". New York, NY: Springer-Verlag, 1997;

[5] G. Rasch, "Probabilistic Models for some Intelligence and Attainment Tests". Copenhagen, Danish Institute for Educational Research, expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago, IL: The University of Chicago Press, 1960/1980;

[6] D.J. Weiss, "The Stratified Adaptive Computerized Ability Test". *Research Report 73-3*. Minnesota, MN: Psychometric Methods Program, Department of Psychology, University of Minnesota, 1973. Retrieved from http://files.eric.ed.gov/fulltext/ED084301.pdf ;

[7] D.J. Weiss, "Strategies of Adaptive Ability Measurement". *Research Report 74-5*. Minnesota, MN:Psychometric Methods Program, Department of Psychology, University of Minnesota, 1974. Retrieved from http://files.eric.ed.gov/fulltext/ED104930.pdf;

[8] N.E. Betz & D.J. Weiss, "An Empirical Study of Computer-administered Two-stage Ability Testing". *Research Report 73-4*. Minnesota, MN: Psychometric Methods Program, Department of Psychology, University of Minnesota, 1973. Retrieved from http://files.eric.ed.gov/fulltext/ED084302.pdf;

[9] N.E. Betz & D.J. Weiss, "Simulation Studies of Two-stage Ability Testing". *Research Report 74-4*. Minnesota, MN: Psychometric Methods Program, Department of Psychology, University of Minnesota, 1974. Retrieved from http://files.eric.ed.gov/fulltext/ED103466.pdf;

[10] K.C. Larkin & D.J. Weiss, "An Empirical Comparison of Two-stage and Pyramidal Adaptive Ability Testing". *Research Report 75-1*. Minnesota, MN: Psychometric Methods Program, Department of Psychology, University of Minnesota, 1975. Retrieved from http://files.eric.ed.gov/fulltext/ED106317.pdf;

[11] F.M. Lord, "The Self-scoring Flexilevel Test". *Educational Testing Service*, Princeton, NJ, 1970. Retrieved from http://files.eric.ed.gov/fulltext/ED042813.pdf;

[12] N.E. Betz & D.J. Weiss, "Empirical and Simulation Studies of Flexilevel Ability Testing". *Research Report 75-3*. Minnesota, MN:Psychometric Methods Program, Department of Psychology, University of Minnesota, 1975. Retrieved from http://files.eric.ed.gov/fulltext/ED111861.pdf;

[13] A. Pyper, M. Lilley, P. Wernick & A. Jefferies, "A Simulation of a Flexilevel Test". The Higher Education Academy, 2014. Retrieved from https://www.heacademy.ac.uk/sites/default/files/comp-256-p_0.pdf;

[14] B.K. Waters, "An Empirical Investigation of the Stratified Adaptive Computerized Testing Model". Minnesota, MN: Air Force Human Resources Laboratory, Applied Psychological Measurement, University of Minnesota Digital Conservancy, 1977. Retrieved from http://iacat.org/sites/default/files/biblio/v01n1p141.pdf;

[15] J.R. McBridge & D.J. Weiss, "Some Properties of a Bayesian Adaptive Ability Testing Strategy". *Research Report 76-1*. Minnesota, MN: Psychometric Methods Program, Department of Psychology, University of Minnesota, 1976. Retrieved from http://files.eric.ed.gov/fulltext/ED121819.pdf;

[16] Y. Zheng & H. Chang, "On-the-fly Assembled Multistage Adaptive Testing". *Applied Psychological Measurement*. Vol. 39(2), 104–118, 2015. Retrieved from http://www.researchgate.net/publication/272369522_On-the-Fly_Assembled_Multistage_Adaptive_Testing.

[17] Y. Zheng, Y. Nozawa, X. Gao & H. Chang, "Multistage Adaptive Testing for a Large-scale Classification Test: Design, Heuristic Assembly, and Comparison with Other Testing Modes". *ACT Research Report Series,* 2012;

[18] G.J. Melican, K. Breithaupt & Y. Zhang. "Designing and Implementing a Multistage Adaptive Test. The Uniform CPA Exam". In the book: W.J. Van der Linden, C.A.W. Glas, "Elements of Adaptive Testing" (pp. 167-189); Springer New York Dordrecht Heidelberg London, 2010. DOI: 10.1007/978-0-387-85461-8;

[19] A. Zenisky, R.K., Hambleton & R.M. Luecht, "Multistage Testing: Issues, Designs, and Research". In the book: W.J. Van der Linden, C.A.W. Glas, "Elements of Adaptive Testing" (pp. 355-372), Springer New York Dordrecht Heidelberg London, 2010.

[20] D. Yan, Ch. Levis & M. Stocking, "Adaptive Testing with Regression Trees in the Presence of Multidimensionality". *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 3, 293-316, 2004. Retrieved from http://www.jstor.org/stable/3701355?seq=1#page_scan_tab_contents;

[21] D. Yan, Ch. Levis & A.A. Davier, "A Tree-based Approach for Multistage Testing", in the Book: "Computerized Multistage Testing. Theory and Applications", Chapter 11 (pp. 169-188). Princeton, NJ: Educational Testing Service, 2014;

[22] G. Makransky & C.A.W. Glas, "An Automatic Online Calibration Design in Adaptive Testing". University of Twente. P 2., 2007. Retrieved from http://www.testpublishers.org/assets/documents/100131Makransky-GlasAutomaticOn.pdf ;

[23] F. Kharshiladze, "The Basics of the Common Analyze". Tbilisi, TSU, (pp. 9-12), 1971 (Georgian);

[24] L. Schwartz, "Mathematical Analysis". Moscow, "MIR" (pp. 21-27), 1972 (Russian).