Scoring the Outcomes of 1-3-3 Multistage Model of Computerized Adaptive Testing

Simon Razmadze
Georgian Computer Society
Tbilisi, Georgia
e-mail: simonrazmadze@gmail.com

https://doi.org/10.62343/csit.2024.16

***Abstract* — The paper presented considers the ordering method of outcome set for multi-stage testing (MST) of 1-3-3 model. The ordering method of outcome set is used for the estimation of results of computerized adaptive testing (CAT). This method is not tied to a specific testing procedure. Acknowledgment of this is its usage for the 1-3-3 model, which is described in the paper. To sort the set of testing outcomes the function-criteria described in the initial article are used here and a comparative analysis of obtained results is performed. The ordered outcome set is estimated by a hundred-point system according to the normal distribution.**

***Keywords* — Computerized adaptive testing, Stradaptive testing, Multistage adaptive testing, Evaluation algorithm, Ordering of a set.**

I. INTRODUCTION

Computerized adaptive testing implies the test adaptation to the level of knowledge of the test. During the testing process the system analyzes the answers and uses them to choose each following question based on the best correspondence to the level of examinee so that the questions gradually become complicated for a well-prepared examinee and simpler for a poorly prepared person. The process of test adaptation for an individual user is mentioned.

This means that the tests must be precalibrated according to their level of difficulty.

The modern Computerized Adaptive Testing (CAT) is based on Item Response Theory (IRT). IRT is a family of mathematical models that describe how people interact with test items [1]. According to this theory test items are described by their characteristics of difficulty and discrimination. Discrimination is independent of difficulty and shows how the probability of a positive response is distributed between different levels of examination. In addition, they can have a so-called “pseudo-guessing” parameter that reflects the probability that an examinee with a very low trait level will correctly answer an item solely by guessing [2].

We will try to create a test assessment system that makes it easy for the test creator to use a computer-adaptive method for creating one’s own test. For this purpose, let us not discuss IRT but another traditional approach to testing - Stradaptive Testing. The term “Stradaptive” is derived from the “Stratified Adaptive”, and it belongs to D.J. Weiss [3, 4].

To express the ordering method of outcomes set, a specific procedure for testing is used in Razmadze et al.’s article [5]. This procedure has an illustrative purpose for the evaluation method. The method described can be used for other similar strategies as well as for multistage testing, one of the models considered in this paper. Similar models were discussed in the articles [6] and [7].

Thus, the paper presented is devoted to the realization of an ordering method of the outcome set, in particular on the example of a three-stage 1-3-3 model.

II. ORDERING METHOD OF OUTCOME SET

The initial article Razmadze et al. [5] considers an original method of CAT result estimation for multistage testing strategy.

The method considers all possible variants of results, which is named an outcome set. The outcome set represents a non-typical unity of different dimensional elements. At Razmadze et al. [5] article comparison criteria for these elements are defined, and principles of ordering of the set are described. The article shows how to receive the final score after ordering the outcome set. The ordered criteria of outcomes set may not be singular; this is confirmed by a comparative review of two examples presented in this work.

Thus, the paper presented is devoted to the realization of an ordering method of the outcome set, in particular on the example of a three-stage 1-3-3 model.

III. THE THREE-STAGE 1-3-3 ADAPTIVE MODEL

A. The scheme of 1-3-3 model

Now let us consider the usage of the ordering of testing result scores in case of multistage adaptive testing. For this purpose, we will discuss the three-stage 1-3-3 model, which is presented in the following scheme [8]:

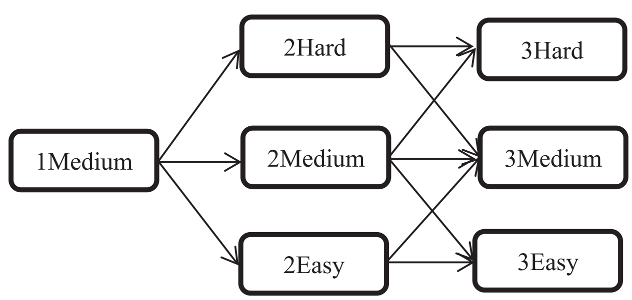

Figure 1. The 1-3-3 MST model

The number indicated in the rectangle of the module corresponds to the stage; the letters correspond to comparative difficulty (Hard; Medium; Easy). Let us number the medium difficulties of modules. Each of these numbers can be considered as the weight of corresponding module item:

Table 1. *The item weights of the three-stage 1-3-3 model*

#	1	2	3	4	5	6	7
Difficulty	3Easy	2Easy	1Medium	2Medium	3Medium	2Hard	3Hard
Weight	1	2	3	3	3	4	5

In the first row of Table 1, all modules are numbered from 1 to 7. We will be using the given numbering for defining the test outcome. Taking into account the complexity levels of the modules, the outcome is expressed as a seven-dimensional vector: $n = \{ c_1, c_2, c_3, c_4, c_5, c_6, c_7 \}$, where c_i represents the number of correct answers of i module, $i = 1, 2, \ldots, 7$. Due to the fact each testee performs only one item on each stage, there can be only 3 components out of a given 7 that are different from 0 in each test outcome. In addition, each c_i component, $i = 1, 2, \ldots, 7$, has a weight, predefined according to Table 1.

For practical reasons, let's assume that the number of questions to be given to the examinee in each module is equal to five. Since only three components of the vector $n = \{ c_1, c_2, c_3, c_4, c_5, c_6, c_7 \}$ can take the whole value from 0 to 5 inclusive (6 options in total), and the other four components are always zero, the total number of test outcomes will be $N = 6^3 = 216$.

The displayed classification can be considered as an analogy to the one used in the item response theory (IRT) (-3; 3) range, where the examinees' abilities are measured [2]. But in this case instead of (-3; 3) range we use the weights provided in Table 1. This does not distort the achievement of the initial task. By considering the weights, the scheme from Figure 1 will transform into the following:

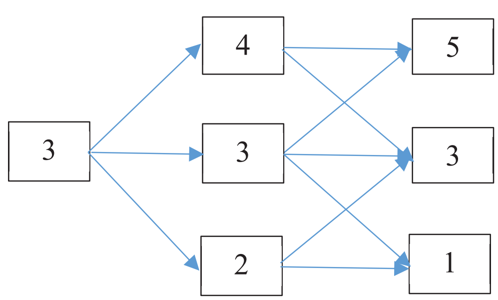

Figure 2. The 1-3-3 MST model with weights

B. *Outcome of 1-3-3 model*

In Razmadze et al. ([5], p. 1656) [article the outcome was defined as a vector drawn from the corresponding numbers of the levels of items obtained during the testing process. In this case, by definition, the outcome vector consists of the components that correspond to the number of correct answers in each module. This is more convenient for using the set ordering method for multistage adaptive tests.

C. *Outcome route*

Modules of the first and the second stage have classification cut-points that define the route of the testing outcome, in other words, choosing the second and third stage modules. Classification cut-point is the number of correct answers within the module that defines the branching – next stage module. Despite where the classification cut-points are chosen the total amount of the testing outcomes is constant and N = 216.

An example discussed in this article on the first stage of 1Medium module has two cut-points: 2 and 4. This means, that in case of less than 2 correct answers (0 or 1) an examinee will be given the easier 2Easy module of the second stage, in case 2 or 3 correct answers - 2Medium module and in case of 4 or 5 correct answers - the more difficult 2Hard module of the second stage.

The second stage modules have the following classification cut-points:

- The classification cut-point of the module 2Easy is 3. If the number of correct answers is less than 3 (0, 1 or 2), the examinee is given the easiest module 3Easy, and if the number of correct answers is 3 or more (3, 4 or 5) – the third stage 3Medium module;
- 2Medium module has two classification cut-points: 2 and 4. This means, that if the number of correct answers is less than 2 (0 or 1), the examinee will be given the easiest module of the third stage 3Easy, if the number of correct answers is 2 or 3 - 3Medium module of the third stage, and if the number of correct answers is 4 or 5 - 3Hard module of the third stage;
- 2Hard module has one cut-point: 3. If the number of correct answers is less than 3 (0, 1 or 2), the examinee is given the 3Medium module of the third stage, and if the number of correct answers is 3, 4 or 5 - the 3Hard module of the third stage.

IV. THE SET ORDERING METHOD FOR SCORING THE OUTCOMES OF 1-3-3 MODEL

A. *Ordering according to the S(n) criterion*

Let us discuss the first criteria from the initial article Razmadze et al. ([5], p. 1658, Formula (4)):

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

where R is a weighted sum of scores of correct answers and M is a weighted sum of scores of incorrect answers.

The corresponding formulas for calculating R and M are given in the article Razmadze et al. ([5], p. 1657, Formulas (1) and (2)). Based on these formulas, in the case of the 1-3-3 MST model, we will obtain the following:

$$R = 3c_1 + 2c_2 + 3c_3 + 4c_4 + c_5 + 3c_6 + 5c_7,$$

$$M = 3d_1 + 2d_2 + 3d_3 + 4d_4 + d_5 + 3d_6 + 5d_7,$$

where c_i is the number of correct answers in i module, and d_i is a number of mistakes in i module, $i = \overline{1,7}$.

The Formula (1), which should be used for outcome estimation, now used in seven-module case. The structure of outcome set of the three-stage model discussed in this article is different from the one discussed in the initial article by Razmadze et al ([5], p. 1656). This means that the domain of a function S(n) is different. Despite this, S(n) function will provide complete ordering of set N in a given case too.

The result is provided in Table 2, where c_1, c_2, c_3, c_4, c_5, c_6, c_7 values are given in the columns B, C, D, E, F, G, H, respectively. The values calculated using Formula (1) are shown in column M. The data is sorted according to M column decreasing order. The table shows the first 10 (left half) and last 10 (right half) testing outcomes' estimation results.

Table 2. *1-3-3 model's outcome estimation by S(n) criterion*

1	N	C1	C2	C3	C4	C5	C6	C7	Normal
2	1	5			5			5	100
3	2	4			5			5	99
4	3	5			4			5	99
5	4	5			5			4	98
6	5	4			4			5	98
7	6	5			3			5	97
8	7	4			5			4	97
9	8	3				5		5	96
10	9	5			4			4	96
11	10	3				4		5	95
208	207	0	3				0		21
209	208	0	0	2					20
210	209	1	0	1					19
211	210	0	1	1					18
212	211	1	1	0					17
213	212	0	2	0					16
214	213	0	0	1					15
215	214	1	0	0					14
216	215	0	1	0					13
217	216	0	0	0					12

B. *Ordering according to the $\boldsymbol{F(n)}$ Criterion*

Let us discuss the second criterion from the initial article Razmadze et al. ([5], p. 1658, Formula (9)):

$$F(n) = R * \frac{A}{\mu}, \quad n \in N, \qquad (2)$$

where R is a weighted sum of scores of correct answers, A is an average complexity of incorrect answers and μ – number of mistakes.

The corresponding formulas for calculating R and A are given in the initial article by Razmadze et al. ([5], p. 1657, Formulas (1) and (3)). Based on these formulas, in the case of the 1-3-3 MST model, we will obtain the following:

$$R = 3c_1 + 2c_2 + 3c_3 + 4c_4 + c_5 + 3c_6 + 5c_7,$$

$$A = \frac{3d_1 + 2d_2 + 3d_3 + 4d_4 + d_5 + 3d_6 + 5d_7}{15 - (c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7)},$$

where d_i is a amount of mistakes in i module, $i = \overline{1,7}$.

$$\mu = 15 - (c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7).$$

The Formula (2), which should be used for outcome estimation, is now used in the seven-module case. The structure of the outcome set of the three-stage model discussed in this article is different from the one discussed in the initial article by Razmadze et al ([5], p. 1656). This means that the domain of a function F(n) is different. Although, it is easy to check that despite this, F(n) function will provide a complete ordering of set N in the given case too.

Table 3. *1-3-3 model's outcome estimation by F(n) criterion*

1	N	C1	C2	C3	C4	C5	C6	C7	Normal
2	1	5			5			5	100
3	2	5			5			4	99
4	3	5			4			5	99
5	4	4			5			5	98
6	5	5			5			3	98
7	6	5			4			4	97
8	7	5			3			5	97
9	8	4			5			4	96
10	9	4			4			5	96
11	10	5			5			2	95
208	207	0	0	2					21
209	208	1	0	1					20
210	209	0	1	1					19
211	210	1	1	0					18
212	211	2		0		0			17
213	212	0	2	0					16
214	213	0	0	1					15
215	214	1	0	0					14
216	215	0	1	0					13
217	216	0	0	0					12

The results obtained by using $F(n)$ criterion is shown in Table 3, where c_1, c_2, c_3, c_4, c_5, c_6, c_7 values are given in the columns B, C, D, E, F, G, H, respectively. The values calculated using Formula (2) are shown in column N. The data is sorted according to N Column decreasing order. Table 3 shows the first 10 (left half) and the last 10 (right half) testing outcomes' estimation results.

The whole table graphically looks as follows (Figure 3):

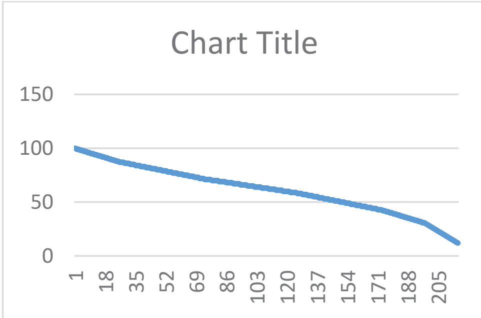

Figure 3. The graph of 1-3-3 MST model testing outcomes' score's normal distribution

V. CONCLUSION

The ordering method of the outcome set can be used in case of different testing procedures. The obvious example of this is the realization of the method for multistage adaptive testing's (MST) 1-3-3 model, which is described in the presented paper.

The author of a test has no direct contact with this method and its specific nuances because the realization of the method is a one-time procedure, carried out during the computerized adaptive testing portal formation.

The method does not require a detailed calibration of the item pool or preliminary testing of examinees to create a calibration sample. The ordering method of outcome set is oriented on the test author; it helps him avoid the problem of preliminary adaptation of test items for the examinee's knowledge level and simplifies the workload at maximum. Preliminary work for the test author might only include the division of test items into several difficulty levels based on expert assessment.

In the situation where there is a lack of information about test items and examinee's level, the method maximally uses the existing information for an examinee estimation: it takes into account all the answers to the questions provided to the examinee and the set of received answers is compared to all the possible variants and placed on corresponding level in the estimation hierarchy.

The paper presents the usage of the ordering method of outcomes set for multistage adaptive testing (MST) model as a sample. The method can be used for different modern testing models, but it is the subject of further research.

REFERENCES

1. Embretson, S. E. & Reise, S. P. (2000). Item response theory for psychologists. Multivariate Applications Books Series. Mahwah, NJ: Lawrence Erlbaum Associates Publishers;
2. Baker, F. (2001). The basics of item response theory. ERIC Clearinghouse on Assessment and Evaluation, Maryland, MD: University of Maryland, College Park. Retrieved from http://echo.edres.org:8080/irt/baker ;
3. Betz, N. & Weiss, D. (1973). An empirical study of computer-administered two-stage ability testing, https://bit.ly/2vjmqou , Technical Report. Research Report 73-4, Psychometric Methods Program, Department of Psychology, University of Minnesota, Minnesota, MN.
4. Betz, N. & Weiss, D. (1974). An empirical study of computer-administered two-stage ability testing, https://bit.ly/2ku9yht. , Technical Report Research Report 74-4, Psychometric Methods Program, Department of Psychology, University of Minnesota, Minnesota, MN.
5. Razmadze, S., Razmadze, M. & Razmadze, T. (2017). The set ordering method for scoring the outcomes of testing in computerized adaptive testing. International Journal of Scientific and Engineering Research (IJSER), Volume 8, Issue 5, May 2017. Retrieved from http://www.ijser.org/onlineResearchPaperViewer.aspx?The-Set-Ordering-Method-for-Scoring-the-Outcomes-of-Testing-in-Computerized-Adaptive-Testing.pdf ;
6. Razmadze, S. The Set Ordering Method for Scoring the Outcomes of 1-2-4 Multistage Model of Computerized Adaptive Testing, Applied Mathematics, Vol. 9 No. 1, 2019, pp. 6-12. doi: 10.5923/j.am.20190901.02. Retrieved from http://article.sapub.org/10.5923.j.am.20190901.02.html ;
7. Razmadze, S. The Set Ordering Method for Scoring the Outcomes of 1-2-3-4 Multistage Model of Computerized Adaptive Testing, LONDON JOURNAL OF RESEARCH IN SCIENCE: NATURAL AND FORMAL. Volume 23, Issue 17, pp. 1-13, „Great Britain Journals Press“, London, 2023. Retrieved from https://journalspress.com/LJRS_Volume23/The-Set-Ordering-Method-for-Scoring-the-Outcomes-of-1-2-3-4-Multistage-Model-of-Computerized-Adaptive-Testing.pdf ;
8. Zheng, Y., Nozawa, Y., Gao, X. & Chang, H. (2012). Multistage adaptive testing for a large-scale classification test: Design, heuristic assembly, and comparison with other testing modes. ACT Research Report Series.