

Adaptive Testing in the System of Learning Outcome Assessment

Ella Hovhannisyan
National Polytechnic
University of Armenia
Yerevan, Armenia e-mail:
hovhannisyanella@gmail.com

ABSTRACT

The paper analyses the existing approaches to the organization of adaptive testing. The methods of adaptive testing are classified. The algorithm of adaptive testing and the method of algorithm development on the basis of AT model are presented.

Keywords

Adaptive testing, AT model algorithm, the method of algorithm development for adaptive testing

1. INTRODUCTION

There are two approaches to testing and assessment of students' knowledge, skills and competences in the educational process. When conducted by the teacher the assessment may be affected by their subjective point of view. To increase the objectivity of the process of evaluation and learning outcome assessment testing may be applied. Testing implies experimental method based on standardized assignments allowing to evaluate the psycho-physiological and personal characteristics, as well as knowledge, skills and competences of the testee. In the course of testing the high level of IT in education enables the process of organization and implementation of students' knowledge monitoring. Testing may serve not only as a method of knowledge assessment and control, but as a tool for current study of the didactic material as a supplement to e-books.

There are different approaches to the building of testing system (knowledge control) for students[1]. To increase the efficiency and reduce the duration of the testing process adaptive testing (AT) is applied. In AT the sequence of assignments is generated directly in the process of testing depending on the actions of the testee. AT implementation supposes the sequence and the number of test assignments which vary for different testees.

2. THE ANALYSES OF THE APPROACHES TO THE ORGANIZATION OF AT

In contrast to the conventional types of testing AT has certain distinct characteristics[2]:

- each testee receives their own set of assignments with different content and length of the test;
- each testee is evaluated individually (according to their level);

The main advantages of AT as compared to the conventional types of testing:

- the possibility to evaluate the abilities of the testee more precisely and with fewer expenses;
- from the point of view of the teacher time is used more effectively, thus there is less negative impact of

additional factors, such as tiredness and inaccuracy, on the results;

- student-teacher direct feedback.

Problems of the organization of computer-based testing are generally viewed from two perspectives: methodical and technical. The methodical one refers to:

- choosing assignments to check the knowledge, skills and competences of the testee;
- planning of the testing process;
- defining requirements for the formation of a set of questions and assignments for the exam, etc.

The technical aspect refers to:

- forming a set of test assignments on the basis of the given approach;
- choosing and implementing the testing parameters;
- choosing the algorithm of the knowledge assessment of testees.

Thus, methods of testing and assessment of the testees' knowledge, skills and competences according to the fulfillment of test assignments are required.

3. AT METHODS ANALYSIS.

To organize AT the following components are necessary to choose and develop:

- methods of test conduct;
- testing result assessment;
- rules of test completion.

3.1. Methods of test conduct

Methods of test conduct differ according to the level of adaptation to individual characteristics of testees and the methods of choosing the parameters for the testing process [3,4].

Partial and full adaptive testing methods are distinguished.

Partial AT supposes that the sequence and number of test assignments is different for progressive, mediocre and weak testees. The choice of the number and complexity of the assignments by the testing system takes into account the testee's answers and/or the level of preparedness of the testee and/or is based on specially elaborated scenarios for the knowledge testing process. For this reason the models of testees are applied. The testee model (TM) may be defined as a set of characteristics of the testee and methods (rules) of its development.

MT should include the following information on:

- the aim of teaching (testing);
- the knowledge of the student within the studied course (the current condition of the educational process);
- the peculiarities of presenting a didactic material and the choice of test assignments and questions (closely related to the knowledge of teaching strategy);
- the rules of change of the testee model according to the results of the work with the student.

Partially adaptive methods are realized by means of random choice taking into account separate parameters of the student model based on the answers of the testee and the model of the didactic material.

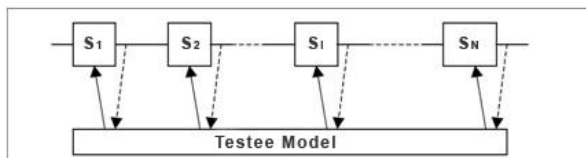
Full adaptive testing allows to implement individual control of knowledge of each testee. Full adaptive methods apply testee models and models of subject spheres.

The methods of testing conduct differ by the choice of the first assignment and by the methods of planning a trajectory for testing [2].

The method of planning a trajectory for testing determines formation of the set of assignments on the basis of the feedback with the testee. Such method allows to check the knowledge individually for each testee corresponding to appropriate level of complexity of the given assignments for the testee and planning individual testing trajectories.

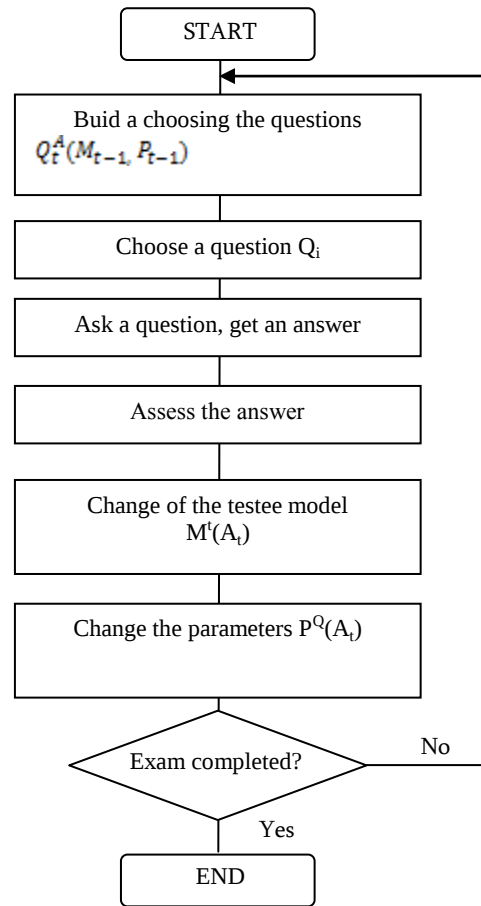
Let us consider the method of testing conduct in depth. On the whole the testing is carried out within a series of sessions S_i , each of which is based on the model of the student. The results of the sessions, in turn, cause (or may cause) the modification of the model (Graph 1). Each session consists of three stages:

- preparation of the assignment for testing (according to the student model);
- student exam;
- evaluation of the exam results and modification of the student model.



Graph 1. Interaction of the testing sessions and testee models.

To form a test assignment a subset of questions is chosen out of a set of questions Q (actual set QA , $QA \supset Q$) to be given to the testee. Originally the set depends on the testee model (M_T) and parameter of questions (P^Q). The set may undergo some changes in the process of the exam. The changes of the actual set based on the answer of the testee (A_T) may be realized through the modification of the student model or through defining other parameters of the questions. The feedback of the actual set and the answer of the student provide the adaptation QA for the student in the course of testing. The general flowchart of testing conduct is the following (Graph 2).



Graph 2. The general flowchart of testing conduct

To monitor the testing process and the model of the student the system should allow the parameters to be introduced to the test. The defining and development of the parameters referring to the description of the testing methods may fall into two categories:

- built-in parameters allow to set the inner algorithms of the testing management;
- user parameters aimed at testing management and forming of the student model are set by the teacher.

By means of built-in parameters it is possible to organize different methods of testing conduct. The following groups of parameters may be distinguished: the type of control, the minimal evaluation, the number of questions, the number of answers, the method of calculation of grades, the scheme of exam conduct, the method of forming a choosing the questions, the method of the first assignment choice, the sequence of assignments (the method of introducing of the assignments), the time of exam conduct.

Let us consider the possible value of these parameters.

The type of control (compulsory parameter)

It decides on the influence of the obtained grades on the future activity of the system. The following values are possible:

Final - intended to test a certain topic. The results of this control decide on whether fulfillment of the assignment is complete.

Intermediary – intended to realize self-testing on a certain topic. The received grade is fixed in the model of the testee, however, it does not affect the fulfillment of the topic itself.

Selective – intended to test the final achievement on several topics. The total grade and the grades on separate topics included in the exam are fixed.

Minimal evaluation (compulsory parameter). It is aimed at the calculation of M_p grade, which ensures the fulfillment of the topic.

The number of questions N_q . It defines the number of questions that the testee should be asked during the exam. The number of questions may be given as:

- a certain natural number;
- a portion of the total number of questions on the given topic – the number in interval $[0,1]$.

The number of answers. It defines the number of answers the testee should give for the topic to consider it covered. $N_a \leq N_q$.

The method of the calculation of the total grade. It defines the method of obtaining the total grades based on the grades for separate questions.

Sum – the total results equals to the sum of grades O_i received for separate answers.

$$O = \sum_{i=1}^{N_q} O_i,$$

Scale – the total results O is calculated according to the scale of minimal and maximal grades.

$$O = \frac{\sum_{i=1}^{N_q} O_i * (max - min) + min}{N_q},$$

Percentage – the total results O is calculated as a percentage of correct answers:

$$O = \frac{m}{N_q} * 100\%,$$

where N_q is the total number of questions asked, and m is the number of correct answers.

The scheme of exam conduct. It conditions the completion of the exam. There are different schemes:

Linear scheme - all the N_q questions are given, afterwards the results are calculated.

Threshold scheme – questions from the sample of N_q are asked till the threshold ‘topic passed’ or ‘topic failed’ is overcome. Thresholds are overcome if the remaining questions cannot change the results of the testing. The threshold ‘topic failed’ is overcome if the maximal results for the remaining questions will not provide the required results M_p . The threshold ‘topic passed’ is overcome if the minimal results of the remaining answers does not result in a grade less than M_p .

Scheme with additional questions. A minimal number of questions N_q is asked, if the total results M_p shows that there is a great deviation in results or the result is nearly critical (M_p), then additional questions are given until:

- the whole list of question on the given topic is over; or
- the maximal allowed number of questions is asked; or
- the deviation in results is reduced to the acceptable level; or

- the threshold of positive results M_p is exceeded. This scheme requires the introduction of additional parameters: N_{max} – the maximally allowed number of questions. ε_p – the level of the result is near to the critical one: the obtained result M is considered close to critical if it meets the conditions:

$$|M - M_p| \leq \varepsilon_p,$$

where ε_p is the level of the allowed deviation of results for example as a mean square deviation:

$$\varepsilon_p \geq \sqrt{\sum_i (m_i - m_{cp})^2},$$

where m_{cp} is the average result for the received answers and m_i is the results for the i -answer.

Scheme with clarification. Writing testing assignments requires a lot of preparation. Questions are asked one by one and the choice of the next question depends on the answer to the previous one and may relate to it. All in all the minimal number of questions N_q is asked. The conditions of completion of the exam are the same as in the previous case. The sample of questions for this scheme should be formed according to the adaptive method.

The method of forming the sample questions. It defines the principle of the inclusion of the questions in the assignment. The following methods of forming the sample may be distinguished: “random”, “by condition”, “proportional to the certain parameter”, “according to the combining parameter”.

“Random” – all questions are selected at random, the same question is not included in the sample twice.

“By condition” - the sample includes only the question for which the function $F(p_1, p_2, \dots, p_n)$ set by the user is true, where p_i is random parameter of the question.

“Proportional to certain parameter p ” - if the complete list of questions on the given topic consists of n_1 questions with the parameter equal to p_1 , of n_2 questions with the parameter equal to p_2 etc., then the sample of questions N_q will include k_i questions such as:

$$k_i = \frac{N_q * n_i}{\sum_{i=1}^L n_i},$$

where L is the number of different P parameters.

“According to combining parameter” – the questions are selected on the basis of certain (main or combining) parameter p_r , according to which they are grouped as to the common concept (term, object, method, etc.) and different degrees of complexity (commonality). The sample is formed to provide that each following question in the group clarifies the previous one. The sequence of questions in the group is determined by another (clarifying) parameter P_y , which may have random semantics. To realize the given scheme the threshold scheme of exam conduct may be applied for each value of the combining parameter.

The method of selection of the first assignment - defines the first assignment which may be set by default (assignment of average difficulty) and be selected individually for each testee or a group of testees based on the available information on the testee/s.

The sequence of assignment - sets the sequence of assignments for the testee. It defines the sequence of completing tasks in the process of testing. From this perspective there are two approaches to the methods of task assignment:

- adaptive testing with constant adaptation;

- adaptive testing with block adaptation.

Adaptive testing with constant adaptation implies that the decision on the change of sequence of assignments is carried out at each stage of testing.

Adaptive testing with block adaptation is a testing where the decision on the sequence of assignments is carried out after the analysis of the results of development of certain set of assignments.

The time of exam conduct defines the allocated time for the completion of the assignment. In case the student does not manage to fulfill the task within the given time, the answers already given are used to evaluate the results of the exam.

It is obvious that the list of parameters through which the users can affect the algorithms of testing conduct is limited. However, it is possible not to limit the number of potential values of these parameters if the user is allowed to decide on the methods of testing.

The setting of parameter values via formulae should be provided. They include arithmetic and logical expressions, consisting of symbols of arithmetic and logical operations, names of built-in functions, names of parameters and values of characteristics which can be calculated in the process of operating the system[1].

3.2 The evaluation of testing results allows to determine the level of knowledge, skills and competences of testees according to the results of the fulfillment of the test assignment, as well as on the completion of the testing process.

3.3 The rules of test completion. The computer based adaptive testing may use the following rules of test completion:

- the required level of preparedness of the testee;
- the number of assignments;
- the time of testing.

The rules of the exam completion define the conditions of testing process completion. A fixed time, a definite number of assignments and a required level of preparedness of the testee are necessary conditions for exam completion. The rules of exam completion may be combined and in their turn result in new rules.

As a criterion of the required level of a testee's preparedness a widely spread rule based on the achievement of the required accuracy of the testee's level of preparedness is used. Limits of the preparedness level of the testee (the minimal and maximal ones) are also applied. The rules of termination such as a fixed time of the exam and a definite number of assignments may also be used, but less often.

4. CLASSIFICATION OF ADAPTIVE TESTING METHODS

As a result of the analysis a set of characteristics defining the means of AT organization was specified: the aim of testing, the means of composing a set of testing assignments, the methods of testing organization, the methods of checking the exam, the methods of evaluation and the rules of exam completion.

The aim of the exam is a significant factor in the process of AT organization, as it defines the content of the exam. The aims of testing fall into two categories: checking the knowledge of a separate topic or section and checking the whole course material.

Depending on the set goal of testing the preparation of the exam material is required, i.e. to work out a method of

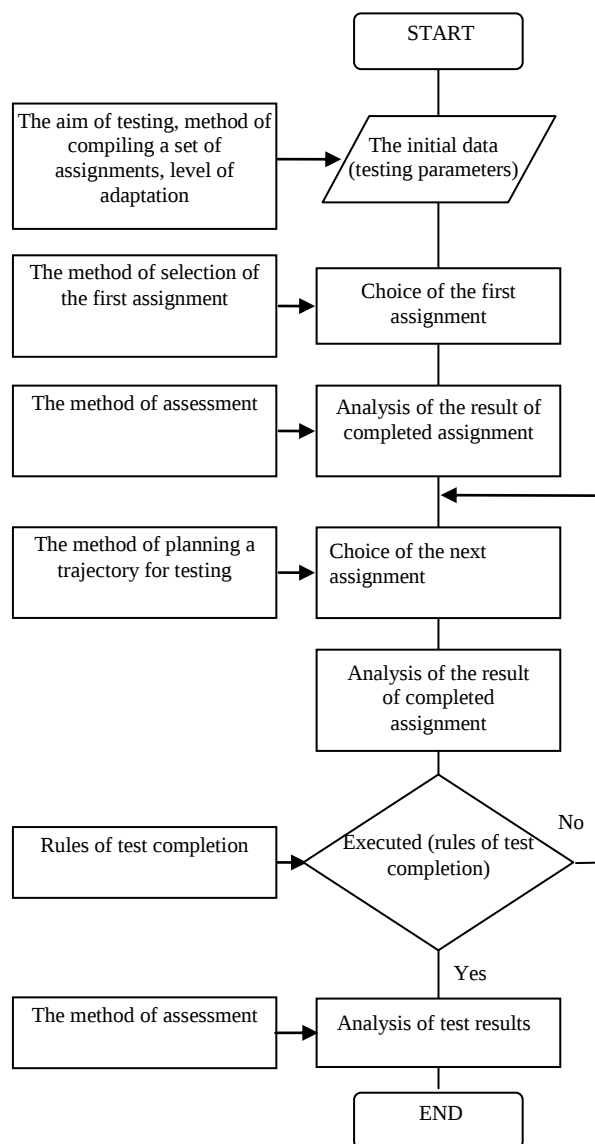
compiling a set of assignments to check the knowledge, competences and skills of the testees.

There are two approaches to the method of compiling assignments:[5]

- thematic approach is a training course consisting of several interconnected sections, where each section contains its own set of assignments;
- task approach is a training course consisting of a set of tasks of different characteristics (type, complexity, etc.)

5. ADAPTIVE TESTING ALGORITHM DEVELOPMENT BASED ON A MODEL

A model based on the approaches to AT organization which defines the principal process of AT and allows to develop AT algorithms which differ by the methods of action realization. The developed classification of AT is offered to be used to choose the method of realization of each activity.



Graph 3. AT model and categories of classification for activity choice.

Graph 3 shows AT model algorithm and categories of classification used to choose the action.

The initial data of the method are the testing parameters:

- parameters characterizing single assignment and its fulfillment (the assignment characteristics; the types of assignment; the time given for task fulfillment; the time allocated to the testee for the test fulfillment)
- parameters used to set AT algorithm (the number of assignments planned for testing; limited time for testing; the aim of testing, characteristics of the testee, etc.)

The choice of the methods of activity realization of AT algorithm based on AT model is made by the developer of the algorithm depending on the aim and conditions of the testing.

6. CONCLUSION

The main advantages of the AT in comparison with classical forms of testing:

- the possibility to evaluate the abilities of the testee more precisely and with fewer expenses;
- from the point of view of the teacher time is used more effectively, thus there is less negative impact of additional factors, such as tiredness and inaccuracy, on the results;
- student-teacher direct feedback.

The analyzed method of AT algorithm is developed on AT model and the classification AT method and may be used by test developers to organize the pedagogical and professional testing. On the basis of the offered method the process of AT algorithm development and the development of the automated system of AT algorithm generation may be automated to meet the requirements.

REFERENCES

- [1] Аванесов В.С. Научные проблемы тестового контроля знаний / Монография. – М.: Исследовательский центр проблем качества подготовки специалистов, 135 с., 1994
- [2] Karpenko A.P., Dobryakov A.A. Model for Automated Training Systems. Overview. Science and Education, № 7-с.3-23, 2011.
- [3] Зайцева, Л. В. Модели и методы адаптивного контроля знаний / Л. В. Зайцева, Н. О. Прокофьева // Educational Technology & Society. – N 7(4), ISSN 1436-4522 (Международный электронный журнал) с.265–277, 2004.
- [4] Березин Н.В. Перспективы создания системы адаптивного тестирования как элемента централизованного тестирования / Научный вестник МГТУ ГА, серия "Информатика", № 38. – с. 26-30, 2001.
- [5] Wim J. van der Linden, Peter J. Pashley Item Selection and Ability Estimation in Adaptive Testing. In Wim J. van der Linden & C. A. W. Glas (Eds.) Computerized Adaptive Testing: Theory and Practice, pp 1-25, 2000.