TESTING OF STUDENTS USING SPECIALIZED CAT MODEL

Svetlana Andjelic[1], Zoran Cekerevac[2], Dejan Andjelic[3]
[1]*Railway College of Vocational Studies, Belgrade*
[2]*Faculty of Industrial Managment in Krusevac, Union University in Belgrade*
[3]*Ministry of Finance of The Republic of Serbia - Tax Administration, Belgrade*
[1]*svetangela@gmail.com,* [2]*zoran.cekerevac@hotmail.com,* [3]*dejan.andjelic@poreskauprava.gov.rs*

Abstract – *The paper presents the basic concept of an original model of a computer adaptive testing (CAT) of students. The test is correlated with the current student's knowledge and provides an opportunity to estimate the optimal number of questions to get a realistic assessment of his/her knowledge. At the end of the paper there are presented some of the results of research carried out using the application based on the CAT model described.*

1. INTRADUCTION

The main questions are: how to test students during a semester, and how to get an objective picture of students' current knowledge on the final exam. The question is also, how to form a grade that will adequately describe student's overall knowledge. It should be borne in mind that one teacher has many students, and, consequently, limited time for monitoring and testing each student. That is why increasing number of teachers resort to electronic tests. As soon as student complete his/her test, according to the number of correct answers, ie. number of points won, the system automatically generates the grade. That way the teacher is exempted from tests review, the accuracy score is higher (human error is avoided), and in a relatively short time it is possible to test more students. However, the problem is: what should be scored and how, in order to obtain a realistic picture of a student's knowledge?

A quality information about student's knowledge can be obtained with a significantly smaller number of questions, if the questions, given to the student, oscilate around the maximum of his current knowledge. Computer-Adaptive Test (CAT) can be a very efficient way of estimation of knowledge. Efficiency is manifested through a reduction in number of questions and reduction of time necessary for testing. The frustration that students experience during the process is also reduced.

2. SPECIALIZED CAT TESTING

The adaptive test is adjusted to the current student's knowledge and it is specific for him/her. Every following question student gets is determined by the wrong or correct answer to the previous question. Such a method of individual testing allows that a very small number of students solve the same test.

CAT process described in this work can be outlined as a process consisting of the following sequences:

1. Entry of the parameters of the model, which are determined through pretesting
2. CAT testing
3. CAT grade formation
4. Evaluation of CAT parameters

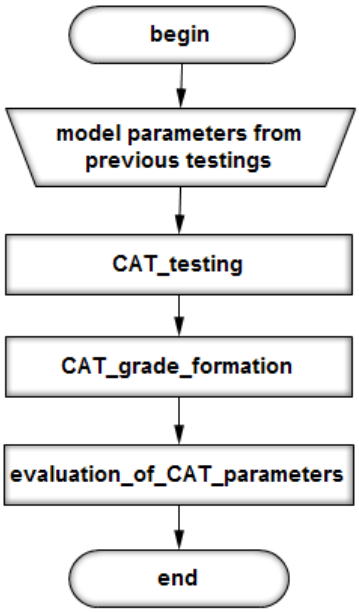

Figure 1 The process of testing a specialized CAT

2.1 STRUCTURE OF THE CAT TEST BY STUDENTS

CAT model contains a test that is divided into different levels of difficulty, where the coefficients of difficulty of each question are expressed in logits. One logit is an equivalent to the difficulty coefficient of 0.73, or 73% probability of the correct answers. This approximation is used in most of CAT models.

Test questions are divided into three levels of difficulty:

I. First, the most difficult, level consists of so called brain teaser (BT), which are the hard questions. Their difficulty coefficient is -2 logit (which equals to value of 0.13 of the difficulty coefficient or 13% probability of student solving the problem).

II. Second level is the medium level and it consists of problems (P), questions (Q) and variations of questions (VQ). Student starts with a problem (-1 logit, or 0.27) and if he/she answers correctly he/she is taken to the first level where gets a brain teaser,

related to the question, but more complex. In the case of wrong answer, he/she gets a theoretical question (0 logit, or 0.50) related to the problem. In the case that student does not give a correct answer, he/she gets a variation to the question (+1 logit, or 0.73), which is the same question, just rephrased to make it easier for the student to understand (since there is a possibility that student did not understend the question properly).

III. Third, the easiest level, consists of basic questions (BQ) which are common for a subject. These questions serve only as a proof that student has some knowledge from the field that is being tested. If the student fell down to this level, he gets, up to two, randomly chosen questions. Coefficient of difficulty of the basic questions is +2 logits, or 0.88.

For each problem, brain teaser, variation of question and basic question, five answers are offered to the student, to reduce the chance of the test passing by random guesses. One of the answers is obviously wrong. Two of the answers contain a mistakes that are not so obvious, but that can be noticed if the student pays some attention to the answer. The remaining two answers are very similar to each other, but only one of them is correct. The other one contains the mistake that students make most frequently. The order of given answers is created randomly and it is changed with each access to a item[1] to reduce the possibility of solving the test on the base of previous experience.

The fields of study that the test covers (set *F_test*), or the overall number of fields *n* ($n = |F\_test|$) are chosen from the list by the teacher. The minimal number of fields that the student needs to cover by answering correctly to questions is also determined by the teacher (*min_nf*). Student answers to *n_stud* fields ($n\_stud = |F\_stud|$). The number *n_stud* depends on convergation of the grade where $min\_nf \le n\_stud \le n$ and $F\_stud \subseteq F\_test$.

The order of fields is given by random choice, without repeating (*random(F_stud)*). This way, it is virtually impossible that two of students get identical questions on the test (Figure 2).

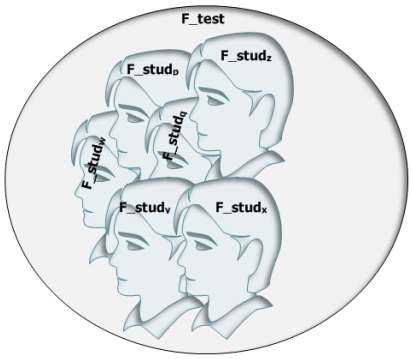

Figure 2 Overlapping of the test by student

[1]Term "item" includes brain teaser, problem, questions, variation of question and basic question

2.2 PROCEDURE OF CAT TESTIG

The algorithm that shows the beginning of the test process, where student selects the starting level (variable level), is shown on Figure 3.

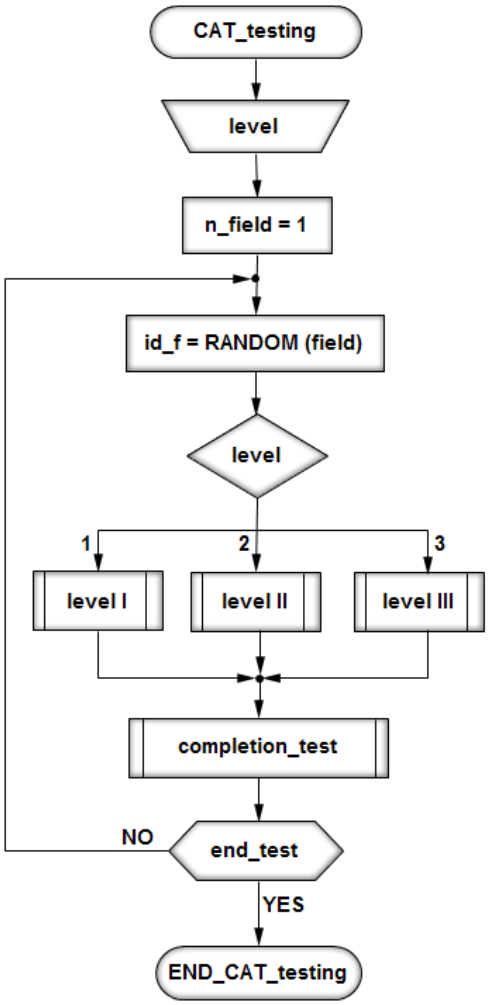

Figure 3 Algorithm of the main program of CAT_testing

At the beginning, the counter field (*n_field*) is initialized to 1 (*n_field*=1). Its value changes in procedure *completion_test*. *id_field* keeps the field code that has been chosen randomly without repetition. Depending on the choices that the student enters, one of the three sub-programs (level I, level II or level III) starts.

The value of *end_test* is determined in the procedure *completion_test*. The condition for the end of the test is defined independently of other parts of the model, with a possible introduction of additional counters. In this way it is possible to change this requirement from test to test, and in doing so is not necessary to change the other algorithms. One of the options for ending of the test might be the maximum number of questions, according to the level, defined by the teacher at the beginning of the test, but it is also possible that all three levels of questions have the same limit. Convergence toward a final grade of the tested field, or total number of fields on the test, could be also criterion for test ending.

2.3 STRUCTURE OF THE CAT TEST BY CONTENT

Each field has its own set of brain teaser, problems, questions, variations of questions and basic questions ($\{BT,P,Q,VQ,BQ\} \in F\_test$). The test itself consists of many items, where $item \in \{BT,P,Q,VQ,BQ\}$, or:

$$item_i = \left\{ \begin{matrix} x_i \mid x_i \in \{BT_{jFi}, P_{jFi}, Q_{jFi}, VQ_{jFi}, BQ_{kFi}\}, \\ i = 1, n\_stud, \quad j = 1, m, \quad k = 1, s \end{matrix} \right\} \tag{1}$$

Where:

m – overall number of subfields for field F_i

s – overall number of basic questions for field F_i and $F_i \in F\_stud \subseteq F\_test$

It should be emphasized that each problem is uniquely connected with brain teaser, question and variation of question, so the mirroring between them is 1:1 (Figure 4).

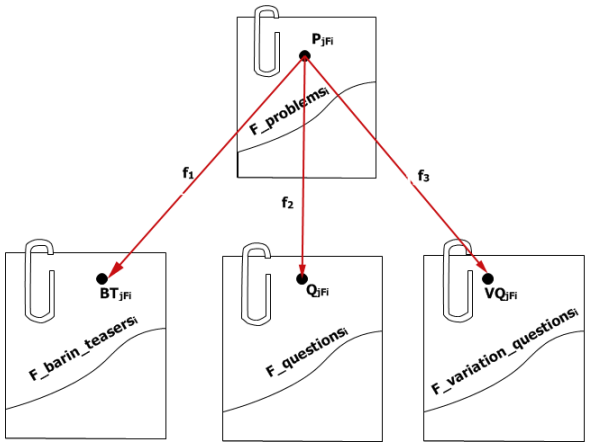

Figure 4 Functions f_1, f_2 and f_3

3. POSSIBLE SCENARIOS OF THE CAT TEST – MATEMATICAL APPROACH

At the beginning of the test student chooses one of three difficulty levels (I, II or III). One of the following two scenarios is possible.

Scenario I – Level I

If the student starts from the level I, he gets a brain teaser from a randomly selected field. The software automatically, by random choice and without repetition generates a field (*random(F_test)*), and after that, also by random choice generates a brain teaser from that field (*random(B)*). In the language of mathematics:

$$item_1 = \left\{ \begin{matrix} BT \mid random(BT_{jF1}) \in F_1, \\ F_1 = random(F\_stud), j = 1, m \end{matrix} \right\} \tag{2}$$

where the variable *m* is already described. If the student answers correctly, he gets another brain teaser, from another field:

$$item_2 = \left\{ \begin{matrix} BT \mid random(BT_{jF2}) \in F_2, \\ F_2 = random(F\_stud / F_1), j = 1, m \end{matrix} \right\} \tag{3}$$

In the case of wrong answer, the student goes to the level II, as well as in the next scenario.

Scenario II – Level II

When starting the test from the level II, the student gets a problem from the random field, by the random choice:

$$item_1 = \left\{ \begin{matrix} P \mid random(P_{jF1}) \in F_1, \\ F_1 = random(F\_stud), j = 1, m \end{matrix} \right\} \tag{4}$$

respecting the notation defined earlier in this text. If the student answers correctly, he goes up to the level one, or follows the first scenario. In the case of wrong answer the student gets a question related to the problem.

$$item_2 = \left\{ \begin{matrix} Q \mid random(Q_{jF1}) \in F_1, \\ F_1 = random(F\_stud), j = 1, m \end{matrix} \right\} \tag{5}$$

Correct answer to the question leads to a new problem, from another field:

$$item_3 = \left\{ \begin{matrix} P \mid random(P_{jF2}) \in F_2, \\ F_2 = random(F\_stud / F_1), j = 1, m \end{matrix} \right\} \tag{6}$$

A variation of the question is given in case of wrong answer:

$$item_3 = \left\{ \begin{matrix} VQ \mid random(VQ_{jF1}) \in F_1, \\ F_1 = random(F\_stud), j = 1, m \end{matrix} \right\} \tag{7}$$

Two situations are possible. A correct answer requires a new problem from another field:

$$item_4 = \left\{ \begin{matrix} P \mid random(P_{jF2}) \in F_2, \\ F_2 = random(F\_stud / F_1), j = 1, m \end{matrix} \right\} \tag{8}$$

This way the Scenario II starts from the beginning.

If the answer is wrong, the student goes down to the lower level and the Scenario III.

Scenario III – Level III

This scenario is used when a student starts with consecutive wrong answers in the Scenario II. The student gets a basic question, randomly chosen from a random field:

$$item_1 = \left\{ \begin{matrix} BQ \mid random(BQ_{jF1}) \in F_1, \\ F_1 = random(F\_stud), j = 1, m \end{matrix} \right\} \tag{9}$$

If the student answers correctly, he gets next basic question from the same field:

$$item_2 = \begin{Bmatrix} BQ \mid random\left(BQ_{jF1}\right) \in F_1, \\ F_1 = random\left(F\_stud\right), j = 1, m \end{Bmatrix} \quad (10)$$

If the student answers correctly to this question, too, the test continues by the Scenario II:

$$item_3 = \begin{Bmatrix} P \mid random\left(P_{jF1}\right) \in F_1, \\ F_1 = random\left(F\_stud\right), j = 1, m \end{Bmatrix} \quad (11)$$

In case of wrong answer to the first or second basic question, it is considered that the student does not have sufficient knowledge in that field (in this case field F_1), and he is taken to another field, which starts from the lowest level:

$$item_3 = \begin{Bmatrix} P \mid random\left(P_{jF1}\right) \in F_2, \\ F_2 = random\left(F\_stud / F_1\right), j = 1, m \end{Bmatrix} \quad (12)$$

Example

A case of a student who started from level I is shown in the Table 1. Further flow of the test developing is based on student's answers, according to the described scenarios. The condition for the test ending is that student answers to 4 fields, out of considered *n*, randomly selected and without repetition.

Table 1. Example of the flow of the test for student X

Serial number of item	Level	id_field	id_item	Answer	Accuracy of the answer
1	1	1	BT11	a	1
2	1	5	BT56	c	0
3	2	5	P56	d	1
4	2	4	P41	b	0
5	2	4	Q41	a	0
6	2	4	VQ41	a	0
7	3	4	BQ47	b	0
8	3	2	BQ23	a	1
9	3	2	BQ27	a	1
10	2	2	P24	c	1
11	1	2	BT24	b	1

Flow of the test, step by step:

1. level = 1, id_field = 1
2. brain teaser (BT11) correctly
3. same level (level = 1), field changes (id_field = 5)
4. brain teaser (BT56) incorrectly
5. level changes (level = 2), same field (id_field = 5)
6. problem (P56) correctly
7. same level (level = 2), field changes (id_field = 4)
8. problem (P41) incorrectly
9. same level (level = 2), same field (id_field = 4)
10. question (Q41) incorrectly
11. same level (level = 2), same field (id_field = 4)
12. variation of question (VQ41) incorrectly
13. level changes (level = 3), same field (id_field = 4)
14. basic question (BQ47) incorrectly
15. same level (level = 3), field changes (id_field = 2)
16. basic question (BQ23) correctly
17. same level (level = 3), same field (id_field = 2)
18. basic question (BQ27) correctly
19. level changes (level = 2), same field (id_field = 2)
20. problem (P24) correctly
21. level changes (level = 1), same field (id_field = 2)
22. brain teaser (BT24) correctly
23. end test

4. RESEARCH RESULTS

Based on the CAT model described, a web application is created in Visual Studio.NET programming environment using C# (C Sharp) programming language. The database was created in MS SQL. The application has been tested in web browsers Internet Explorer and Mozilla Firefox.

The test was conducted on a sample of 100 students at the Faculty of Industrial Management in Krusevac. Students were first tested on the classical manner (pencil and paper, a total of 14 issues), and then they made the test by using a computer and CAT application.

According to the survey, 82% of students declared that the conventional way of testing was not convinient to them (Figure 5). Their responses can be summarized as follows:

- test is not adapted to the student (35%),
- students are reluctant to answer questions that require long answers (23%),
- for the same answer to a question in many cases students scored different number of points (25%).

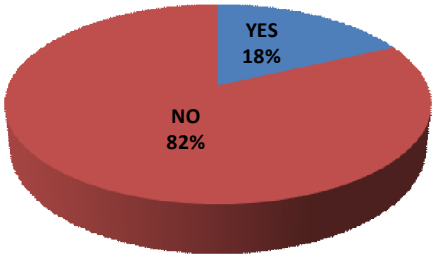

Figure 5 Do you prefere classic way of testing?

Results of the survey revealed that students show a great confidence in the CAT testing. Students need a little preparation, both psychological and in terms of using the software. At the beginning they were a little suspicious and scared, but after short training, psychological problem was overcome.

Even 70% of students answered that CAT testing suit to them, with the following explanations:

- students have declared that they got neither too easy nor too difficult questions; based on their conclusions, one can get the impression that the questions were adapted to each student (38% of students),
- the testing is interesting and not dull (32%),
- it is easier to answer questions that have already offered answers (45%),
- teacher does not affect the score (20%).

The motivation of students is higher. Namely, in the classic test, good students are required to answer some trivial questions, that are, for them, "tedious", and, on the contrary, the weaker students respond to complex and difficult issues, that are also the "tedious". It is evident that both groups during the time, when respond to inappropriate questions, loose the concentration, and, also, loose the motivation.

Using CAT model student was asked to select one of three levels of difficulty to start the test. On that way, an excellent student gets the opportunity to answer correctly to different brain teasers, very quickly finish the testing and get high grades. Analogously, the weaker students are able to respond to questions that by the weight correspond to their current level of knowledge and get the appropriate grades. The important fact is that every student has a subjective feeling that a test was designed specifically for his/her level of knowledge, because the test adapts to him/her, and ask questions that are "close", or that are not much above or much below of student's level of knowledge. It provides further, that the student is maximally motivated and focused in an effort to solve the test correctly.

74% of students consider that, the applied CAT testing reflects their knowledge objectively (Figure 6). Students, who have declared that they do not like CAT tests, were belonged mostly to the group of students that declared their poor use of the computers, mostly the students of older ages (between 30 and 50).

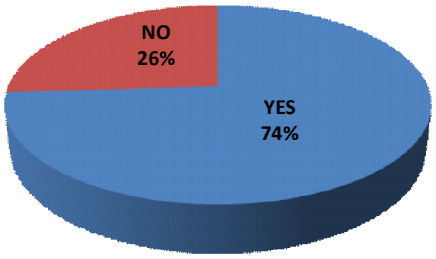

Figure 6 Does the CAT test applied objectively reflects your knowledge?

Students cited the following reasons: lack of familiarity with computers, even the fear of computers (10% of students), they were not accustomed to the interface (5%), they felt a lack of independent verification (and self evaluation) of their answers, and earned grades (15%).

5. CONCLUSIONS

CAT system estimates student's level of knowledge by the smallest number of questions. On the basis of previous answers, the system selects the next question, whose answer gives the best information about the the level of student's knowledge. Presented model of personalized testing might be considered as a form of the simulation of the "oral examination" of students, all in order to achieve more realistic, more objective, images of the level of adoption of specific teaching materials and the level of student's capabilities.

The presented model and CAT test application are flexible and allow further research. With proper refining and upgrading they can lead to the high level expert systems and the artificial intelligence technology. That is the ultimate goal.

LITERATURE

[1] Andjelic, S. *''A supplement to objective evaluation of student work using computer adaptive testing''*, PhD thesis, Defended 10/08/2010, Faculty of Industrial Managment, Union University in Belgrade

[2] Andjelkovic J. *''Knowledge estimation by the use of computer adaptive testing''*, In Proceedings of 9th International Symposium SymOrg, Zlatibor, 2004.

[3] Desmarias, M. C. & Pu, X. *''A Bayesian Inference Adaptive Testing Framework and Its Comparison with Item Response Theory''*, International Journal of Artifical Intelligence in Education, 15, pp. 291-323, 2005.

[4] Garrido Morro, C. *''Different approaches to computer adaptive testing applications''*, Retrieved, Martch 15, 2010. http://cs.joensuu.fi/pages/whamalai/sciwri/cristina.pdf

[5] Guzman, E. & Conejo, R. *''A model for student knowledge diagnosis through adaptive testing''*, Proceedings of 7th International Conference on Intelligent Tutoring Systems, pp. 12-21, 2004.

[6] Ju Gin-Fon, N. & Bork, A. *''The implementation of an adaptive test on the computer''*, In Proceedings of 5th International Conference on Advanced Learning Technologies (ICALT), pp. 822-823, 2005.

[7] Tao, Y.-H., Wu, Y.-L. & Chang, H.-Y., *''A Practical Computer Adaptive Testing Model for Small-Scale Scenarios''*, Educational Technology & Society, 11(3), pp. 259-274, 2008.

[8] Van der Linden, W. J. & Veldkamp, B. P. *''Constraining item exposure in computerized adaptive testing with shadow tests''*, Journal of Educational and Behavioral Statistics No 29, pp. 273-291, 2004.

[9] Vehkalahti K., *''Reliability of measurement scales''*, Finnish Statistical Society Statistical Research Reports 17, 2000., Retrieved, December 20, 2008. http://ethesis.helsinki.fi/julkaisut/val/tilas/vk/vehkalahti/reliabil.pdf

[10] Wise S. L., Kingsbury G. G., *''Practical issues in developing and maintaining a computerized adaptive testing program''*, Psicological, 21, pp. 135-155, 2000.