



Auto-calibrated Online Evaluation: Database Design and Implementation

Lorentz JÄNTSCHI¹, Sorana-Daniela BOLBOACĂ²

¹*Technical University of Cluj-Napoca, 15 Constantin Daicoviciu, 400020 Cluj-Napoca, Romania;* ²*“Iuliu Hațieganu” University of Medicine and Pharmacy, 13 Emil Isac, 400023 Cluj-Napoca, Romania*
lori@academicdirect.org, sbolboaca@umfcluj.ro

Abstract

An auto-calibrated online evaluation system for student knowledge assessment was developed and has been implemented at the Faculty of Materials Science and Engineering, the Technical University of Cluj-Napoca. The system is available for students' end-of-course evaluation on physical chemistry, microbiology and toxicology, and materials chemistry topics and it can be used at the test center. The auto-calibrated online evaluation methodology, databases and program implementation, computer-system design and evaluation are reported.

Keywords

Auto-calibrated online evaluation, Computer-system design, Multiple choice questions (MCQs), Evaluation methodology, Databases and program implementation

Introduction

Educational measurement at the university level has been moved in the last years from the paper-and-pencil testing towards the use of computer- and/or Internet-based testing. Computer-based testing refers to performing examinations via stand alone or network

computers [1]. The examination questions are displayed on a computer screen and the student choose their answers using the computer's mouse or keyboard; the responses are recorded by the computer.

Computer-based tests can be found at all educational levels (pre-university [2] and university [3]) and in many universities all over the world (Lagos state University College of Medicine, Nigeria [4]; Heriot-Watt University, Edinburgh, Scotland [5]; University of Hertfordshire, United Kingdom [6, 7]; Loughborough University, Leicestershire, U. K. [8]; Pennsylvania State University, U. S. A. [9]; University of Toronto, Ontario, Canada [10]; University of Southern Queensland, Queensland, Australia [11]; University of Hawai'i at Manoa [12]; Kulak University, Belgium [13]). Many researchers compared the equivalence of computer-based and paper-and-pen tests and most of them conclude that computer may be use in many traditional multiple-choice test settings without any significance on student performance [14, 15, 16]. Some differences were observed on tests which include extensive reading passages, when the students proved to obtained lower performance on computer-based tests comparing with on paper tests [17, 18].

Starting from the experiences obtained by creation of the multiple-choice examination system for general chemistry topic [19, 20], the present paper concerned with issues involved in creation of an auto-calibrated online system, computer-system designs, database and program implementation, and system evaluation.

Material and Method

Multiple choice questions methodology

An important part of a computer-based testing is represented by creation of the multiple-choice questions (MCQs), which contain the procedure of creation, storing and managing of MCQs.

The anatomy of multiple-choice test contains two basic parts: a statement or a situation, a problem (*question* or *stem*) and a list of suggested solutions (*alternatives* or *options*). The stem may be construct in form of a question or of an incomplete statement and the list of options must always contains at least one correct or best alternative one and a number of incorrect options (*distractors*). The distractors must appear as *plausible* solutions

to the problem for the students who have not achieved the interest objective measured by the test and *implausible* solutions for those students who have achieved the objective (only the correct option(s) should appear plausible to these students).

Regarding the anatomy of the questions, each problem has five suggested options and must test comprehension, application as well as students' analysis abilities. Regarding the number of correct options, the multiple choice questions cover the followings classes:

- **Single correct answer:** all except one of the options are incorrect, the remaining option is the correct answer;
- **Best-answer:** the alternatives differ in their degree of correctness. Some options may be completely incorrect and some incompletely incorrect, but at least one option is completely correct;
- **Multiple responses:** two or more of the options but never all five options are keyed as correct answer. These kind of questions can be scored in different ways, but most frequently use are:
 - *All-or-none rule:* one point if all the correct answers and none of the distractors are selected, and zero points otherwise. This was the alternative choused for our system;
 - *Scoring each alternative independently:* one point for each correct answer chosen and one point for each distractor not chosen.

Methodology of scores and final mark

In order to provide an objective evaluation the scores and final mark methodologies take into consideration multiple factors some of them directly related with students knowledge and some related with associated activities.

There were implemented two types of scores which are directly related with students acquired knowledge: the score of an individual test and average scores of student tests. The score of an individual test characteristics and modality of computing are presented in table 1.

In order to compute the final mark, since each student had the possibility of testing his/her knowledge as many time as he/she want in the imposed period of examination, individual average score is computed. In computing of the individual average score, all student scores were taken into consideration with one exception. If a student had at least two scores, the lowest one is deleted and the left scores are used to compute his/her average testing score.

Table 1. Scores and mark methodologies

Parameter	Description
tmn_rc_global	The average time needed to give a correct answer (take into consideration all tests stored into database)
tm_rc	The average time needed to give a correct answer for the current test (just the data from the current test are taken into consideration)
C_t	The time coefficient Formula: $C_t = tm_rc / tmn_rc_global$
C_{c-a}	The coefficient of correct-answered questions Formula: $C_{c-a} = nr_rc / nr_rc_global$
m_g	The geometric mean Formula: $m_g = \sqrt{C_t * C_{c-a}}$
$T_{score} = m_g * 10$	The test score Formula: $T_{score} = m_g * 10$

The final mark can take values from 4 (the lowest average of test scores; the exam is failed) to 10 (the higher average of test scores). The final mark methodology took into consideration the students scores as well as his/her accomplish (bonus, up to 2-3 points) or non-accomplished (penalty, subtracted up to 1 point) of assumed activities (e.g. stems banking construction activities). There are applied penalties (subtracted 0.5 point) every time when the student gives up after beginning of a test.

Examination methodology

The computer-based examination methodology imposes the followings:

- Place of examination: at test center, C414 room, the Technical University of Cluj-Napoca;
- Type of examination: computer- and teacher-assisted. This modality was chosen in order to avoid cheating and copying. The teacher was responsible for login of each student to the computer-based testing interface;
- Period and time of examination: imposed by the academic year structure (period of examination; one month) and by the teacher (time of examination; were defined for each day from the week the hours when the lab was open for computer-based test examination);
- Number of MCQs: thirty;
- MCQs test generation: each test was automated generated by the application, using a double randomization method. The first randomization was applied in choosing the stems for each test and the second one for displaying the options order. The second randomization assured the randomization of correct options (it is known that there is a tendency to make “B”, “C”, and/or “D” the correct choices and to avoid “A” and/or “D” [21]);

- Choosing the correct answers: by mouse selection;
- Number of computer-based tests: as many as the students want;
- Test results: auto-calibrated of scores and final mark according with data stored into database every time when a student performed a new test.

Database design

The implementation of computer-based testing system was performed by the use of a relational database *chemistry* that store multiple choice banking (*..._tests*), students information (*..._users*) and information regarding the evaluations (*mcqs_results*). The MySQL database server [22] stores the database on vl.academicdirect.ro server, AcademicDirect domain.

The structure of the table that store the multiple choice banking (*..._tests*) is in table 2. In table 2 are specified the name of the field, its type and description of the field.

*Table 2. The structure of the *..._tests* table*

Field name	Field type	Specification
<i>`id`</i>	bigint(20)	An auto increment value which is manipulated by database server and serves as primary key for record unique identification
<i>`name`</i>	varchar(30)	Store the name of the user which create the steam
<i>`item`</i>	varchar(255)	The steam
<i>`o1`</i>	varchar(255)	First possible option
<i>`o2`</i>	varchar(255)	Second possible option
<i>`o3`</i>	varchar(255)	Third possible option
<i>`o4`</i>	varchar(255)	Forth possible option
<i>`o5`</i>	varchar(255)	Fifth possible option
<i>`a1`</i>	tinyint(4)	Correctness (1) or incorrectness (0) of the first possible option
<i>`a2`</i>	tinyint(4)	Correctness (1) or incorrectness (0) of the second possible option
<i>`a3`</i>	tinyint(4)	Correctness (1) or incorrectness (0) of the third possible option
<i>`a4`</i>	tinyint(4)	Correctness (1) or incorrectness (0) of the forth possible option
<i>`a5`</i>	tinyint(4)	Correctness (1) or incorrectness (0) of the fifth possible option

The structure of the table that store the students information (*..._users*) is in table 3.

*Table 3. The structure of the *..._users* table*

Field name	Field type	Specification
<i>`id`</i>	tinyint(4)	Auto increment value that serves as primary key for record identification
<i>`name`</i>	varchar(20)	Store user second and first name
<i>`p`</i>	varchar(255)	Store the penalties apply to the student (if were apply)
<i>`bonus`</i>	float	Store the bonus points (if were apply)
<i>`pass`</i>	varchar(32)	Store encrypted password

The structure of the table that store the students information (`mcqs_results``) is in table 4.

The `..._tests`` table are related with `..._users`` and respectively `mcqe`` tables.

Table 4. The structure of the `mcqs_results`` table

Field name	Field type	Specification
<code>`id`</code>	bigint(20)	Auto increment value that serves as primary key for record identification
<code>`subj`</code>	vvarchar(10)	Store the topic abbreviation (there were included into the system MQCs for three topics: physical chemistry (<i>phys_chem</i>), microbiology and toxicology (<i>microb tox</i>), and materials chemistry (<i>mate_chem</i>))
<code>`name`</code>	vvarchar(30)	Store the names of the student that performed the test
<code>`suid`</code>	vvarchar(32)	Store encrypted password
<code>`qlist`</code>	vvarchar(255)	Store the list of the questions that appeared on the test
<code>`tb`</code>	vvarchar(17)	Store the dates (dd.mm.yy hh.mm.ss format) corresponding with the start of examination
<code>`te`</code>	vvarchar(17)	Store the dates (dd.mm.yy hh.mm.ss format) corresponding with the end of examination
<code>`p`</code>	tinyint(4)	Store the number of obtained points (the number of correct answers)

System interface

PHP (Hypertext Preprocessor) [23] was use to implement the auto-calibrated online evaluation system. The characteristics of the main programs are in Table 5.

Table 5. Auto-calibrated online evaluation environment: programs characteristics

Program name	Function	Remarks
<code>index.php</code>	Main window of auto-calibrated online evaluation system	The main interface contains a description of the computer-bases test environment, of scores and respectively final mark methodologies
<code>wrong_questions.php</code>	Display the wrong questions stored into database	There are display the name of the author, the stem, the options and the correctness or incorrectness of each option
<code>index_add.php</code>	Allows submission of a new MCQ and/or correction of a MCQ stored into database	The interface allows: • to go to <code>new_user.php</code> interface; • to add new stems and options and to define the options correctness or incorrectness; • displaying of previous saved MCQs. The changing of the stems and associated options is password protected (just the user which create the MCQ are allowed to change it); • to link towards special symbols and/or characters.
<code>new_user.php</code>	Allows registration of a new user	
<code>tests.php</code>	Testing interface	The access to this interface is password protected (just the teacher as administrator had the password)
<code>test_points.php</code>	Display scores and final mark	
<code>definitions.php</code>	Securities specifications	

Results and Discussions

The auto-calibrated online evaluation system was created and it is available via the address: http://vl.academicdirect.org/general_chemistry. Until now, the system was use in order to test students' knowledge at the Faculty of Materials Science and Engineering, the Technical University of Cluj-Napoca, Romania regarding following topics:

- Physical Chemistry: third year of study;
- Microbiology and Toxicology: forth year of study;
- Materials Chemistry: first year of study.

The access to the system is restricted by checking the IP addresses, being available just at the test center.

MCQs banking

Three MCQs banking were created, one for each above presented topics. The total numbers of MCQs stored into databases according with each topic are: 424 for Physical Chemistry topic, 363 for Microbiology and Toxicology topic and 865 for Materials Chemistry topic.

The distribution of the stems with one, two, three and for correct options according with each topic, express as relative frequency (f_r (%)) and associated 95% confidence interval (95% CI) compute by the use of online software which use the Newcomb method with continuity correction [24] are in Table 6.

Table 6. MCQs banking characteristics according with topics

Correct options	Physical Chemistry		Microbiology and Toxicology		Materials Chemistry	
	f_r (%)	95%CI	f_r (%)	95%CI	f_r (%)	95%CI
One	49.3	[44.4, 54.2]	65.3	[60.1, 70.1]	34.4	[31.3, 37.7]
Two	26.9	[22.8, 31.4]	16.3	[12.7, 20.5]	23.2	[20.5, 26.2]
Three	16.5	[13.2, 20.5]	10.5	[7.6, 14.2]	20.9	[18.3, 23.8]
Four	7.31	[5.1, 10.3]	7.99	[5.5, 11.4]	21.4	[18.7, 24.3]
Total	100		100		100	

Analyzing the results from Table 6 it can be observed that the number of MCQs with one correct option is almost fifty percent on physical chemistry topic and exceed the half from the total number of questions for microbiology and toxicology topic (65.3%). The heterogeneous distribution of the questions with one, two, three and respectively four correct

options could be a disadvantage of the MCQs banking on physical chemistry and microbiology and toxicology topics. Looking at the MCQs banking on materials chemistry topic it can be observed that there is a homogeneous distribution of questions with one, two, three, and respectively four correct answers. Thus, the proportion of MCQs with one correct answer is a little bit greater comparing with the proportions of MCQs with two, three and four correct answers, but the differences are significant.

What can be done in the future regarding the MCQs banking? On physical chemistry, respectively microbiology and toxicology topics, through creation of another questions, the distribution of the items with two, three and four correct options can become homogeneous.

Computer-assisted tests environment

The main interfaces of auto-calibrated online evaluation environment for all three topics are similar and look like in figure 1.

<p>☉ Hot links:</p> <p>Wrong Questions</p> <p>Add/Modify Questions</p> <p>New Student</p> <p>Testing</p> <p>Obtained Points</p>	<p>Auto-calibrated online evaluation is developing to perform on computers from C414 lab. A student can test his/her knowledge however desire. The database contains more than 300 problems. Each test contains 30 problems with five possible options. A problem is consider to be correct solve when were exactly marked the correct options. Each correctly solved problem bring 1 point. The test is against-time. The moment when the test is generated and the moment when the test solutions are sent to database are recorded. It is not possible to give up the test after it was generated. Every desertion is penalized when the final mark is computed. Testing mark calculation: Calculation of tests average: ... Mark calculation: - The lowest average of test is associated with mark 4 (four); - The highest average of test is associated with mark 10 (ten); - The mark is given by the placement of the test mark between the lowest and highest test average. Final mark calculation: ... Comments? lori@academicdirect.org</p>
---	---

Figure 1. The main interface of the auto-calibrated online evaluation system

The system allows management of MCQs on predefined topics. In order to introduce a MCQ into database, the user must to register and to choose a password. When the user decide to introduce MCQs into database, by choosing `Add/Modify Questions` link the system

allows to choose the user name by selecting it from a drop-down list and open the interface of adding new questions and/or modifying previous saved questions (see figure 2). In modifying interface, first all MCQs created by the user are displayed and by selecting the stem 'id' which the user want to change it, the user can change the selected stem, associated options and definition of options correctness or incorrectness. A modification of a steam and/or associated options is password protected, just the user which created and saved the question having the right to change it.

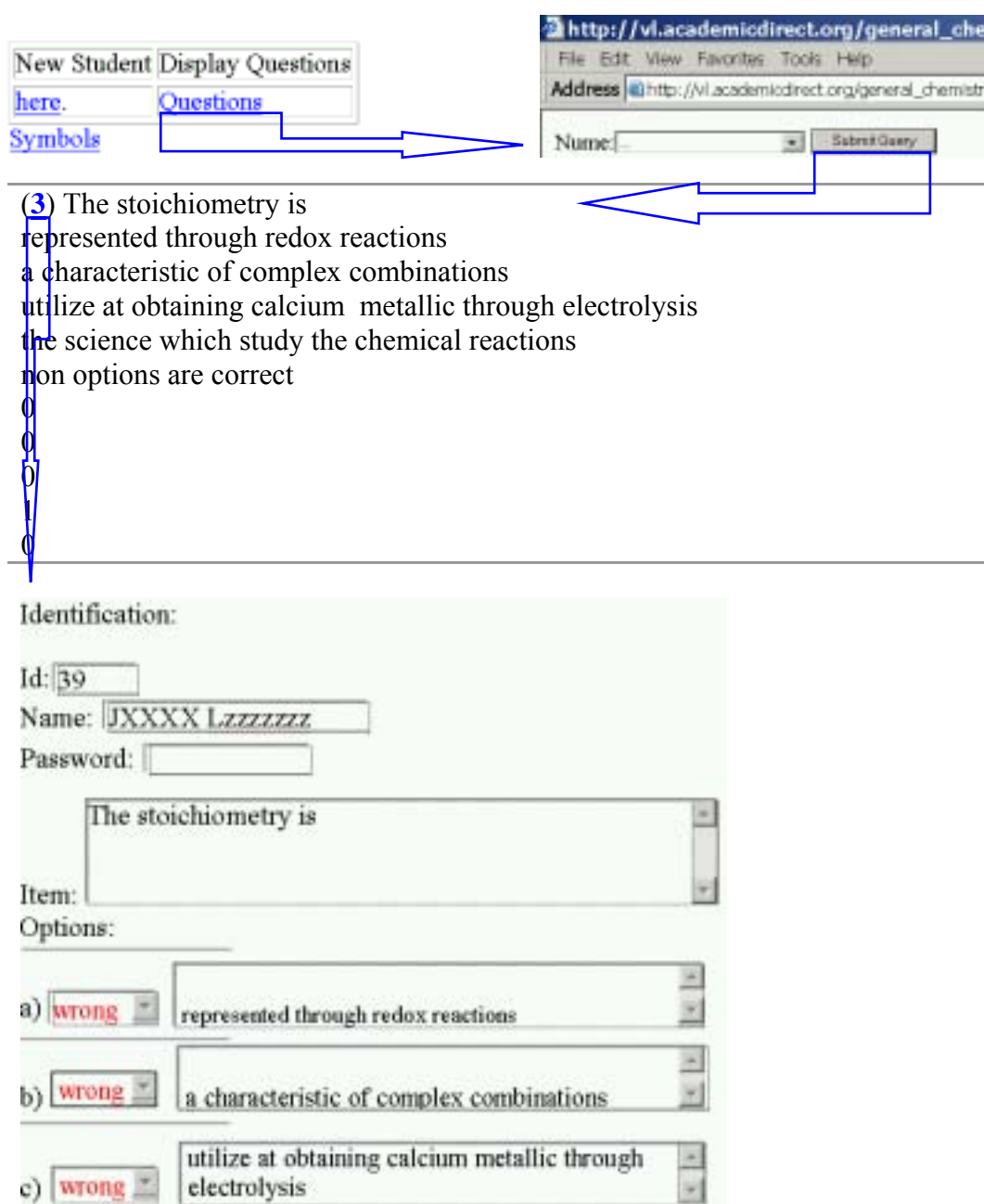


Figure 2. Interface for modifying of a MCQ

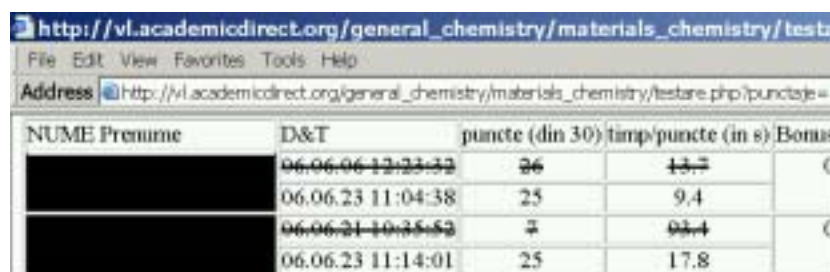
The testing is imposed to take place at the test center and the access to it is protected by teacher password. The teacher is the one which open the test for each student. For each student, the system display automated on the screen thirty MCQs, each student having a different set of questions. Each steam has five possible options (the number of correct options is not specified). On the left side of each option, there is a radio-button that allows selection of the correct answers.

The finalization of the test by selecting of the 'I finished the test' button did not allow returning at the testing interface and display on the screen the identification data of the student (first and second name), the time when the test was started and when was ended and the number of correct answers from a maximum of thirty.

From the main interface of the system there is access to the obtained points and marks of all students which performed the exam on one of the three specified topics ('Obtained Points' link, figure 1). The results are display as data organized in tables and graphical representations (distribution of the obtained scores and distribution of the final marks).

As results, the data are organized into three tables:

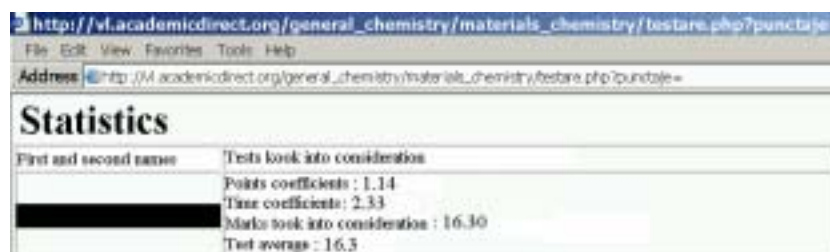
- First table: contains student identification details, date and time of the examination, the number of obtained points, the time necessary to give a correct answer, and the bonus points when was applicable (see figure 2);



NUME Prenume	D&T	puncte (din 30)	timp/puncte (in s)	Bonus
[blacked out]	06.06.06 12:23:32	26	13.7	0
[blacked out]	06.06.23 11:04:38	25	9.4	
[blacked out]	06.06.21 10:35:52	7	93.4	0
[blacked out]	06.06.23 11:14:01	25	17.8	

Figure 2. General results of computer-base testing

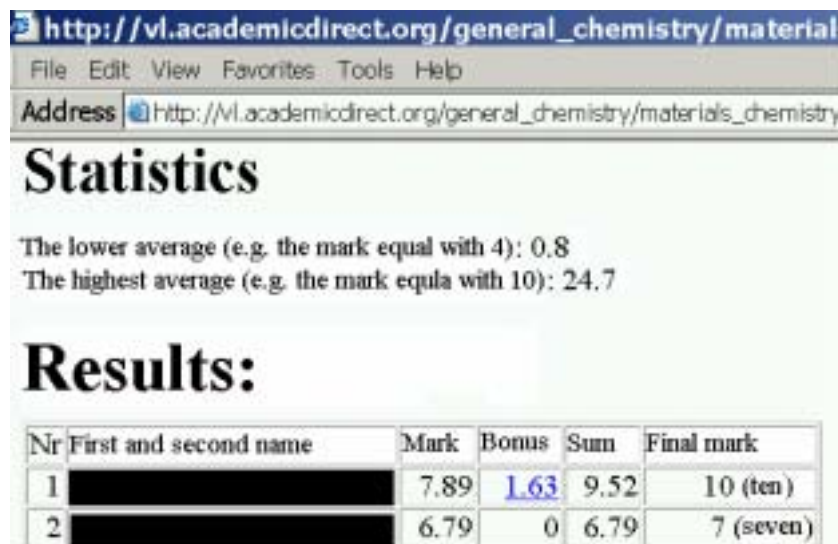
- Second table: contains details regarding marks and averages of testing with specification of the average time needed to give a correct answer and the average of the correct answers. (see figure 3);



Statistics	
First and second names	Tests took into consideration
[blacked out]	Points coefficients : 1.14
[blacked out]	Time coefficients : 2.33
[blacked out]	Marks took into consideration : 16.30
[blacked out]	Test averages : 16.3

Figure 3. Detailed statistics of tests results

- Third table: contains final statistics with specification of the correspondence between the points and the mark equal with 4 and between the points and the mark equal with 10 and the final results (student identification data, final score, the bonus points, and the final mark) (see figure 3).



Nr	First and second name	Mark	Bonus	Sum	Final mark
1	[REDACTED]	7.89	1.63	9.52	10 (ten)
2	[REDACTED]	6.79	0	6.79	7 (seven)

Figure 4. Final results

The advantages offered by the developed online evaluation system are:

- Questions and their options are carefully structured. In these conditions the students could not guess the correct answers by applying a process of elimination;
- Questions have one or up to four correct options and the number of correct answer(s) is not specified;
- Each student has a different individual test. In this condition, the problem of copying among students is eliminated. Randomization of items and of associated options prevent learning of the correct options associated to an item (for example: if the student learned that for item no. 412 the correct options are “A”, “C”, and “D” it is sure that on his/her test the order of the options will not be the same) and promote learning of knowledge;
- The analysis of each student is performed through an auto-calibrated approach each time when a student performed the test, being automatically, in real time. Thus, at the end of examination each student can see his/her performances.

The system has also its disadvantages. Being an online evaluation system required access to computers and to a local-network. The system also required as prerequisite from the students' minimal knowledge of using a computer. In order to enable students to familiarize with the new evaluation environment, before the final examination, pre-exam evaluations

were open to be performed at the examination center by any interested students and as many times as they wanted.

Conclusion

The auto-calibrated online evaluation system is a valid and reliable solution in students' knowledge evaluation, providing a multi criteria objective assessment environment.

Acknowledgement

Research was partly supported by UEFISCSU Romania through the project ET46/2006.

References

-
- [1] Olsen J. B. [online], *Guidelines for computer-based testing*. Retrieved May/June 2000. Available at <http://www.isoc.org/oti/printversions/0500olsen.html>.
- [2] Ananda A. L., Gunashingam H., Hoe K. Y., Toh Y. F., *Design of an intelligent on-line examination system*, Computers & Education, 13(1), 1989, p. 45-52.
- [3] Frosini G., Lazzerini B., Marcelloni F., *Performing automatic exams*, Computers & Education, 31(3), 1998, p. 281-300.
- [4] Oyebola D. D., Adewoye O. E., Iyaniwura J. O., Alada A. R., Fasanmade A. A., Raji Y., *A comparative study of students' performance in preclinical physiology assessed by multiple choice and short essay questions*, African journal of medicine and medical sciences, 29(3-4), 2000, p. 201-205.
- [5] Beevers C. E., McGuire G. R., Stirling G., Wild D. G., *Mathematical ability assessed by computer*, Computers & Education, 25(3), 1995, p. 123-132.



-
- [6] Lilley M., Barker T., Britton C., *The development and evaluation of a software prototype for computer-adaptive testing*, Computers & Education, 43(1-2), 2004, p. 109-123.
- [7] Barker L. M., Britton C. T., *Automated Feedback for a Computer-Adaptive Test: A Case Study*, Proceedings for 9th CAA Conference 2005 [online] [cited 13 May 2006], Available at http://www.caaconference.com/pastConferences/2005/proceedings/LilleyM_BarkerT_BrittonC.pdf
- [8] Croft A. C., Danson M., Dawson B. R., Ward J. P., *Experiences of using computer assisted assessment in engineering mathematics*, Computers & Education, 37(1), 2001, p. 53-66.
- [9] Ariew R., *A diagnostic test for students entering a computer-assisted learning curriculum in French*, Computers & Education, 3(4), 1979, p. 331-333.
- [10] Brydges R., Classen R., Larmer J., Xeroulis G., Dubrowski A., *Computer-assisted assessment of one-handed knot tying skills performed within various contexts: a construct validity study*, The American Journal of Surgery, 192(1), 2006, p. 109-113.
- [11] Taylor A. J., *Self Test: a flexible self assessment package for distance and other learners*, Computers & Education, 31(3), 1998, p. 319-328.
- [12] Brown J. D., *Computers in language testing: present research and some future directions*, Language Learning & Technology, 1(1), 1997, p. 44-59.
- [13] Demeersseman E., Wylin B., Panen J., *TOETS: development of a computer assisted assessment system*, Proceedings for 5th CAA Conference 2001 [online] [cited 13 May 2006] Available at <http://www.caaconference.com/pastConferences/2001/proceedings/n1.pdf>
- [14] Peak P., *Recent Trends in Comparability Studies* [online]; Released August 2005 [cited July 2006] [thirty-four pages], Available at http://www.pearsonedmeasurement.com/downloads/research/RR_05_05.pdf
- [15] Russell M., Goldberg A, O'Connor K., *Computer-Based testing and Validity: A Look Back and Into the Future* [online], Released August 2003 [cited July 2006], Available at: <http://www.bc.edu/research/intasc/PDF/ComputerBasedValidity.pdf>

- [16] MacCann R., *The equivalence of online and traditional testing for different subpopulations and item types*, British Journal of Educational Technology, 37(1), 2006, p. 79-91.
- [17] Mazzeo J., Harvey A. L., *The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature (ETS RR-88-21)*, Princeton, NJ: Educational Testing Service, 1988.
- [18] O'Malley K. J., Kirkpatrick R., Sherwood W., Burdick H. J., Hsieh M.C., Sanford E.E., *Comparability of a Paper Based and Computer Based Reading Test in Early Elementary Grades*, Paper presented at the AERA Division D Graduate Student Seminar, Montreal, Canada, April, 2005.
- [19] Naşcu H. I., Jäntschi L., *Multiple Choice Examination System 1. Database Design and Implementation for General Chemistry*, Leonardo Journal of Sciences, 5, 2004, p. 18-33.
- [20] Naşcu H. I., Jäntschi L., *Multiple Choice Examination System 2. Online Quizzes for General Chemistry*, Leonardo Electronic Journal of Practices and Technologies, 5, 2004, p. 26-36.
- [21] Atkinson M., *Writing Effective Multiple Choice Exams*, The Successful Professor 1(6), 2002, p. 2-4.
- [22] MySQL [online], MySQL AB, ©1995-2006 [cited 2006 June], Available at <http://www.mysql.com>
- [23] PHP [online], The PHP Group, ©2001-2006 [cited 2006 June], Available at <http://www.php.net>.
- [24] Newcombe R. G., *Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods*, Statistics in Medicine, 17, 1998, p. 857-872.