

## Infrastructure for e-Testing

Goce S. Armenski and Marjan Gušev

**Abstract:** The characteristics of the society in which we live, where knowledge and the ways of its use are the most important in everyday life, brings new challenges for higher education. The extensive use of technology in learning and working is forcing its use in the assessment process. A lot of software packages exist in the market to realize automated assessment. In this article we analyze different methods used for testing and present new frontiers, especially in cases where the number of students is very big (several hundreds), and in cases in which students can take exams every month. Using the results from this analysis we have designed and developed a new assessment system. We also give a report of the results from using e-testing tool for assessment of student knowledge, concentrating on the effectiveness of its use for assessment purposes.

**Keywords:** e-Testing, eLearning, web-based assessment.

### 1 Introduction

In the past two decades lots of changes happened in the society where people live. These changes which affected peoples' lives were mostly a result of economical globalization, increased meaning of knowledge, and information and communication revolution.

The growth of global economy creates unusual competitive business relations. Today's companies are not just competing with their local concurrent, but they are forced to expand their businesses on foreign markets, where they are faced with bigger competition. In order to survive in this highly competitive marketplace, they need to possess a capability to locate, understand and use the possibilities which arises every day.

---

Manuscript received February 2, 2004

The authors are with Faculty of Natural Sciences and Mathematics, Arhimedova 5, 1000 Skopje, University of Ss. Cyril and Methodius, Skopje, Macedonia (e-mail: [armenski, marjan]@ii.edu.mk).

During 80's there was evolution from "industrial society", where natural resources and manufacturing were main force for moving economies and societies forward, into "information society", where main focus is given to information. During 90's it was concluded that the information itself can not bring significant changes, which resulted in "knowledge society", in which the most important is the way in which the information is transferred into knowledge, and the way that knowledge is used. According to [1] knowledge can be defined as productive use of information. In a business sense, knowledge is an ability to react in profitable and productive way.

In [2] Alvin Toffler presents new theory and strategies for adaptation in these new social changes. Many of his predictions are towards the role of the education in these adaptation processes. One of his most cited prediction is redefining the education in this way:

"The illiterate of the 21st century will not be those who cannot read and write, but those who cannot learn, unlearn, and relearn"

High education institutions have the main role in the process of redefining the models for acquiring knowledge and skills, and the models for obtaining additional resources. The traditional ways of giving lectures, delivery of materials, and knowledge assessment can not be adapted completely to this new situation. In the processes of giving lectures and delivery of materials, audio-visual and web technologies can be used as a solution, but the teachers have limited options in the process of assessing student knowledge.

The problem is much bigger in courses attended by a large group of students. Usually teachers look for help from their colleagues in the process of conducting the assessment and checking the student work, a solution which have its own abnormalities in subjective marking, and in late feedback delivery to students.

There are lots of systems for automatic assessment on the market. Some of them exist as independent software packages for computer based assessment, web based assessment or electronic assessment, but also there are systems which are part of some already existing systems for distance learning.

The system for electronic testing at the University "Ss Cyril and Methodius" is a result of 30 month development of concepts and software, which will be used for conducting assessments very often on which more that 150 students will take part. The original idea was based to create a system that can help realization of the exams for a number of students bigger than 150 and realize the study concept when each student is allowed to apply for an exam each month. Afterwards this idea was expanded to cover more principles to realize an independent system of testing with a lot of intelligence. The final aim of this software is to make an electronic testing system applicable both for conventional and distance learning, to be implemented for quizzes or exams for any purpose. Later the possibilities of this system were

used to improve the learning process.

In this paper we analyze the concepts which are common for all automated assessment systems, and the technologies used for their creation. Using the results from this analysis, we will define a new model of a system for electronic testing. Using this system we will give answer to the question: how the use of technology influences the assessment process?

## 2 Methodological aspects

The process of acquiring knowledge, skills and information is known as learning. Lots of scientists define learning as organization of the behavior based on experience. Existence of large number of theories about the way people learn, implicates existence of large number of definitions about learning. Common for all of them is existence of internal connection between individuals' mind, the rest of the nerve system and the environment in which that individual is [3].

From business point of view, learning is defined as a tool for achieving certain goals, which can be: increase of the working force effectiveness, cost decrease, productivity improvement, quality improvement of the products and services etc.

Learning is interactive process in which the environment where learning takes place, the learners, teachers and the tools for learning are involved. The interaction between the students and teachers is characterized with flow of information in a form of dialog, book or computer software.

### *Theories for learning*

The ways people acquire new knowledge was a subject of research many years, and as a result, few theories that define the process of learning and training exist. Two most commonly used are instructivism and constructivism.

The instructivism as a theory is based on behavioristics theory (in which the process of learning is compared with a black box, where what enters and what comes out does matter, and what is happening inside the box is considered as not important) is dominant theory in the first two thirds of 20-th century, and is a base for many education systems in the world. According to this theory the knowledge exists out of the students mind, and is transferred through training. "According to this theory, learning flows in a mostly unidirectional part, proceeding from the knowledgeable authority (teacher), or from instructional content, to the passive learner." [4].

If the main goal of every training is personal education and intellectual improvement of the individual, which will bring benefit to the society, the instructivism is not useful neither for the student, nor for the teacher or the society.

During 20th century, there was change in learning and training theories [5]. It

was very obvious that those students which are more interested in what they learn are giving more energy in learning some specific curricula and are achieving better results. This fact, and the rapid development of the technology, helped in creation of one of the mostly used and accepted theories about learning and training process at the moment, in which “more attention is given to the learning and the support of the students while they learn how to create their understanding about culture and society which part they are” [6] [7] - theory of constructivism. According to this theory, knowledge does not exist out of students mind, but “it is actively created by the student, and can not be acquired passively by the environment” [8]. Compared to instructivism, constructivism is more concerned with learning instead of training, and puts the student in the center of the learning process, dealing mostly with its individual skills. Previous knowledge which every student possesses influences on processing and transformation of new knowledge through strategies like analysis, decision making, interpretation and selection. This way the student is capable of thinking further from the information he possess [9].

Constructivism as a theory is widely used in the past years, but still, it is criticized because of some insufficiency. Most often used critics is concerned with the issue that according this theory every person can create his own picture of the world, which is equally important as any other picture some other person created, which basically means that the world exists only in the mind of the person which actually created it. Thinking in this way, it is very hard to standardize knowledge, and at the same time to assess that knowledge. Also one of the other problems is that every person tends to “discover the same things which are learned from the humankind with centuries ago” [5].

#### *Assessment*

Assessment takes central place in the learning process. The assessment nature have deep impact on the way people learn [10], defining the contents they will assume as important and defining the way they will spend their time [11]. The most of the students think that the assessment is the most important activity in learning process, becoming obsessed with the results from those assessments.

Assessment is a process in which examples of person's attitude are taken at particular time and they are evaluated. According to the evaluation of these examples, conclusions are made for the person's achievement, potential, intelligence, attitude or motivation. Different forms of assessment exist and each of them has different use. Mostly used form of assessment is the one used in the education institutions and is defined as “continuous assessment of student knowledge by their teachers held where learning takes place” [12]. According [13], the assessment is “obligational or optional activity or exercise where one explicit intent is to assess student progress or learning achievement in a unit of study”, and is considered as “an integral part of the teaching and learning process”.

*Functions of assessment*

The impact which assessment has on the learning process is maybe the mostly used reason for its use. Waters and McCracken [14] points out that the assessment should be used as a tool to improve learning, instead of using it as separate activity, or as a tool to mark the students and rang them. Improper use of the assessment process is taking students away from the learning goals, because very often their goal is to achieve better results on the assessments.

The improvement of the learning process through assessment, is one of the segments on which lots of attention is paid while designing learning systems today. For example Keys concluded that there is 14% improvement when the tests are given every week instead every month [15]. Pikunas and Mazzota also mention improvement of 10% when the tests are given every week instead every 6 weeks [15]. The ideal learning system would be the one which will be able to adapt according to the needs and abilities of the students. Often assessment use and analysis of the feedback, makes possible to display content specialized for the particular student. Learning system designed in this way will be capable to lead the student and help him in achieving the goal, which is learning the given concepts.

*Assessment automation*

Through the history there is very small change in the way the assessment process is held [16], but in the past decade because of social [17] and technical reasons [18], significant change can be seen. In the literature many attention is given on assessment automation by using technology in its process. Bennett [19] writes that the innovations in technology will significantly influence the assessment process. Research in the same field has been done by many other researchers from different science areas. In their research papers they are giving different reasons for automation of the assessment process, which can be categorized as pedagogical and practical. Bull summarizes those reasons and says that “the use of computers in the assessment process can improve student learning, the assessment process itself, and at the same time will give significant advantages for the academic workers” [20]. Summons says: “...with large numbers of students with different learning styles, there is a need to develop instructional arrangements that maximize student learning, while trying to minimize the cost in terms of time, effort and money” [21]. Maybe one of the most important reasons for technology implementation in the assessment process is the fact that it is becoming a tool which is used in the process of learning and working. The forum for Education and technology [22] suggest that “...as technology is integrated in the curricula, the way that assessments should use the same tools used in the learning process”. Incompatibility between the way of learning and assessment, the use of medium different than the one used in the learning process, may give wrong results for the achievements. [23] [24].

As it is often case when implementing technology for automation of any pro-

cess, necessary research should be conducted to locate the existing procedures in order to define the best ways for its implementation. The technology implementation itself, does not give any guaranties that the new systems will be accepted and effective. There is a long list of failed projects because the consequences of using the technology were underestimated, and because the environment failed to accept those changes [25].

#### *Terms in automated assessment*

The broadest term which is used in literature when discussing assessment automation is computer assisted assessment. This term covers any use of computers in the process of assessing knowledge, skills and abilities of individuals. The term encompasses the use of computers to deliver, mark and analyze assignments or examinations. It also includes the collation and analysis of data gathered from optical mark readers (OMRs) [26].

One of the main parts of computer assisted assessment is computer-based assessment which is defined according [27] as: questions and assignments delivered to the student using computer or other electronic device which can be integral part of any network or can exist alone. In most cases the student is entering its responses with keyboard and those are saved and marked electronic. In [27] and in [28] there is very similar definition about computer-based testing. Good definition about term testing is given in [29], as technology that includes facilities to assist in the making up of practice quizzes, tests, exams, and other assignments. We use this definition and further we conclude that computer-based testing is a method for testing by using computer hardware and software. As a follow up we define the electronic testing systems as dynamic computer systems that support computer based testing. They include various methods for data base management and artificial intelligence to support decision-making process for generation of testing material. All these activities are realized through WEB based Internet technologies. Some other authors give definitions and differences of testing as computer-based, PC-based, WEB-based etc. In this article the term computer based testing will be used when dealing with any testing realized by computer. Under PC based testing we will consider use of computer for testing purposes which is not connected in network with other computers, and under Web or Online testing we will consider use of a computer for testing which is connected in any kind of network with other computers and which uses testing system realized using Web technologies.

When the testing is Web based, the access to the testing system can be done not just using a computer but also using any electronic device (mobile or static) which have access to Internet. This fact brings us to define the term electronic testing which includes use of any electronic devices in the process of testing. The system which accepts access from different electronic devices is called electronic testing system.

### 3 Concepts of computer based assessment systems

#### *Question bank*

The core of every testing system are the questions it has. Under question bank we will understand a database of unique questions, which have necessary parameters for simple selection during test construction. The necessary parameters can be: the question weight, the area where it is categorized etc. Every system should have a module for question entry, a module for importing questions from other existing question banks and a module for removing existing questions without ruining its concept.

Additional energy is invested in the development of few standard forms of structuring the elements from question banks, in order to provide compatibility among different systems. There exists few standards which are still in development, but the most promising is the IMS system [30]. The specifications defined in this system are given using XML. As a result of this development, exchange of question banks among the systems for computer based testing can be expected.

Testing systems consist of different types of questions. They can be divided in two categories [31]: fixed response questions and free response questions. In literature, fixed response questions are also called objective questions, while free response questions - unobjective questions. The main difference between these types of questions is the nature of the answer.

Most of the systems for automated assessment use fixed response questions [31], because these questions give possibility for automatic evaluation. The characteristics of the questions which are categorized here are that they have text with which they describe the problem and a list of possible answers from which student can choose the one he thinks is the true one. The types of questions categorized in this category are:

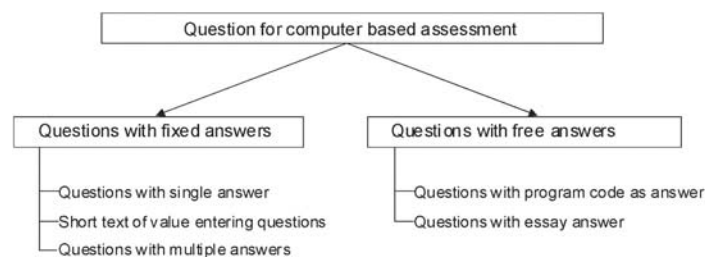


Fig. 1. Clasification of the questions for computer based assessment.

*multiple answers questions* - most used type of questions in automated assessment. Variations of this type of questions are: true-false questions, questions with

priority setting;

*short answer questions* - type of questions where as answer should be provided short text or numerical value which was discovered as a result from some computation;

*hot-spot questions* - type of questions which have good graphical surrounding and are characterised with a need from a student to identify object or a position, to transfer elements from one place on the screen to another or to connect some elements.

A main characteristic of free response questions is that there is no predefined answer. They are usually used when assessing higher levels of Bloom taxonomy. The types of questions which can be categorized as free response questions are questions with program code as answer; questions with essay answer.

Questions used in question banks can be categorized also depending on the way the user answers the question. According to this classification, they can be divided in two categories. In the first category are those questions where the answer is chosen from few possible options using the mouse, and in the second category are those questions where the answer is entered using the keyboard in a form of text. Graphical view of this classification is given on Figure 2.

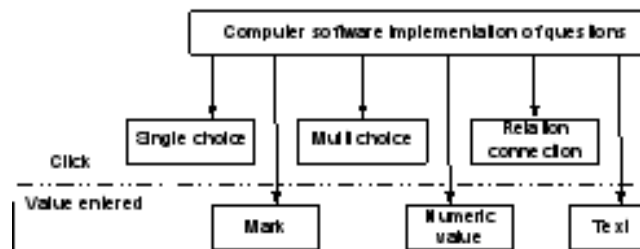


Fig. 2. Computer software program implementation can be classified according to the input.

Use of question banks has lots advantages over static creation of tests. The advantages are mainly concerning the speed in which tests are created, and in possibility to choose the learning objects from which the student will be assessed. The statistics which are given for every question independently, can give better information about the student competence comparing the statistics which are given for the whole test, where it is very hard to determine where exactly the problem is and which learning object is less accepted by the students.

#### *Test creation algorithm*

The question bank is the core of every testing system. The way those questions are selected from the question bank and the way they are presented to the student is called test creation algorithm. Test creation algorithms are very dependent of the



test delivery models. There are few test delivery models among which main difference is the level of adaptation to the characteristics to the person which knowledge is tested. Using that criteria we can differentiate 5 different models for test delivery [32] showed on Figure 3.

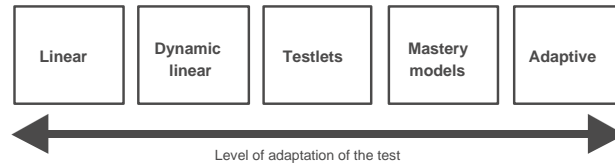


Fig. 3. Models for test delivery [32].

*Linear tests* are not adaptive to users and consist of predefined questions and predefined order. They present traditional testing by using computers. Assessment can be done automatically and results be summarized..

*Adaptive tests* depend on student's knowledge. The parameters of test generation are defined dynamically during the test according to given answers.

#### *Display systems*

The use of complex algorithms for test creation, and the use of complex questions are dependent of the technology which is used (network speed, server characteristics, the characteristics of the screen etc). There exist lots of configurable parameters which can have influence on the results from any testing. Differences in visualization (text size, the chosen font, the color, the contrast), audio presentation (sound quality), and the speed of displaying the data are some of the reasons which can have effects on the results [33]. One of the key parameters which have influence on the reading speed and on the understanding what is displayed is the screen resolution [33], because of its influence of the amount of information displayed on the screen.

#### *Marking and results reporting*

One of the main advantages of automated assessment is the possibility to display the results of the assessment immediately after the assessment. Different testing systems implement different approaches towards marking and results displaying. All of these approaches tend to achieve easy understanding and interpretation of what is displayed [34]. The process of marking should take place at the moment when the answer of any question is definitely entered. Depending of the characteristics of the system concerning the possibilities to change the entered answers, the evaluation of the entered answers can be after it is entered or at the end of the test.

According to the concepts for evaluation of the answers, we can differentiate two different concepts for displaying the results from the testing. If the evaluation of the answers is made right after they are entered, the results can be shown at that

time too, but if the evaluation is made after test completion, than the results are displayed at the end too.

Automated evaluation of the answers is possible if the tests are constructed of fixed answer questions. That is why this type of questions is usually used when using automated assessment. One disadvantage of the objective tests (tests consists of fixed answer questions) is the possibility to achieve good results using intelligent guessing, or by lucky guessing instead of demonstration of knowledge. The research showed that this disadvantage can not be completely eliminated, but the effectiveness of the students which use these methods can be lowered with increasing test length [35]. The results from these researches [35], show that if the test consists of 50 questions, the percentage of the students which can pass the test with guessing is 1.4%, while if the test consists 60 questions, the percent is 0.6%.

Trying to make a difference between those students which know the answer, those which admit that they do not know the answer, and those who guess, some systems are trying to lower the guessing influence by implementing weight of the entered answer [36] [37]. When students enter answer to some question, they are supposed to provide information how much they are sure that the selected answer is the right one.

Maybe one of the most effective ways to eliminate guessing is the implementation of corrective points, usually known as negative marking. When this approach is taken, the student gets negative points every time he chooses wrong answer. There are lots of academic discussions about negative marking, in which some researchers have negative attitude towards it, but also there are researchers which strongly support this kind of marking. Some researchers think that well prepared tests, which have questions with different weight and cover wide part of course curricula, can minimize the influence of guessing [38]. The use of negative marking can have negative influence to those whose knowledge is assessed, because they will be not able to show partial knowledge, and also can increase the time used to take the test, because the students will be more careful in answering the questions.

The automated evaluation of free text answers is hot topic in the past few years. While some researchers are trying to make systems which evaluate these answers in some way, most of the others think that it is almost impossible to build a system like that. The answers theoretically can be written in thousands of ways, using different combinations of words and sentences. The most of these words have tens ways in which they can be used in the sentences in different contexts.

The most popular systems for automated evaluation of essay answers are: Project Essay Grade (PEG), Intelligent Essay Assessor (IEA), E-rater, Bayesian Essay Test Scoring System (BETSY). Every one of these systems has its own specific way for evaluation and marking of the answers..

## 4 eTest software description

### System architecture

The system is realized as WEB application with three layer architecture. Main reason to realize this system not as classic client server application is the characteristic of Web applications to be installed on one computer and used on any computer on Internet through common web browser. This is very convenient for students since they are not obliged to come at university to use the system, and can efficiently use e-business paradigm anytime, anywhere. A good survey of Internet as Business Infrastructure and also for e-Testing system can be found in [29].

The three layered architecture of the system is shown on Fig. 4, realized by separate database layer, application layer for basic system modules and user interface layer.

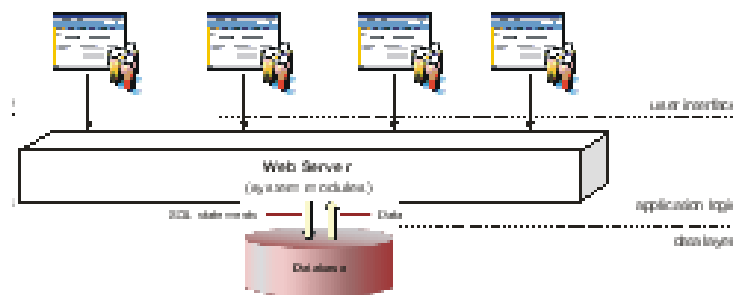


Fig. 4. Three layered architecture of the system.

User interface layer is connected with the application logic layer and is used by the users which submits requests. Any computer on which web browser is installed can be used as a client, and it is not depended on the company which produced the browser. Also any wireless device can be used which has integrated web browser in it.

The application logic layer consists of web server, on which files coded in ASP (Active Server Pages) scripting language are stored. The operating system is Windows 2000 server, with Internet Information Server installed. The function of this layer is to accept the requests from the user interface layer, generate answer to those requests and send them to the users.

Data layer consists of SQL Server 2000 database. This database is in communication with the web server on which dynamic web pages are executed, and where the answers to user requests are generated.

### Course organization

The basic structure of the system consists of courses which material is divided in lectures. A tree like organization of lectures is implemented. Each lecture con-

sists of smaller parts and each part consists of different sets and finally of learning objectives, as shown in Fig. 5 [39]. The course material in the lesson is divided in at least three parts (in Fig. 5 marked as A, B and C). For each part there are at least 4 sets of questions. Each set of questions consists of at least 5 questions. At least one of these questions is hidden and is intended for final test exam, not for testing purposes. The remaining questions are candidates for online testing.

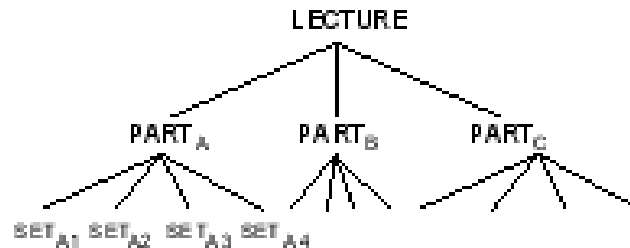


Fig. 5. Tree like organization of the lectures.

The system for eTesting can have unlimited number of courses which are administrated by users with special privileges to administrate courses. The courses are independent from each other and have their own structure and question bank.

#### *Question types*

The idea for creating a system which will be able dynamically to create tests, lead us to use a model with question bank from which the questions will be selected. The questions are divided in different learning objects, according to the course organization described above. The number of questions available for every course is very big (usually 1500 questions per course).

The electronic testing system realized as computer-based testing is realized in such a way that each test generated will measure verbal, quantitative and analytical skills related to a specific field of course study. A different time constraint and score mark is associated to each area. The area consists of a set of questions defining one concept or one knowledge skill. We differ three classes of questions: verbal, quantitative and analytical questions, similar to the description in [40].

A testing area that consists of verbal questions defines a specific concept or definition. 2 points are usually associated to be gained for a correct answer if the question is answered in 60 seconds. The verbal measure tests the ability to analyze relationships among component parts of sentences, to analyze and evaluate written material and synthesize information obtained from it, and recognize relationships between words and concepts.

A specific quantitative measure is expected to be calculated, or more sophisticated concept to be explained in the case of quantitative questions. 4 or 6 points are usually associated to be gained for maximum of 240 seconds. Basic skills and un-

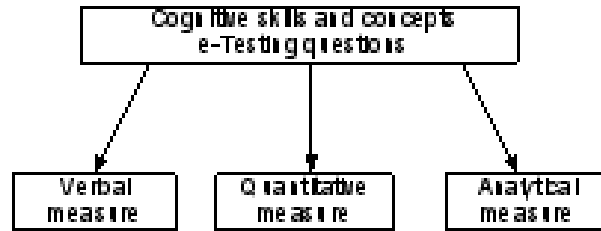


Fig. 6. Knowledge based cognitive skills and concepts can implement e-Testing models by three types of questions with: verbal, quantitative and analytical measure.

Understanding of elementary concepts, as well as the ability to reason quantitatively and solve problems in a quantitative setting or to explain more sophisticated concepts are tested in the quantitative measure tests.

Rather complicated formula is used to calculate the expected answer; or rather difficult concept has to be explained for analytical questions. We usually award up to 10 points for a correct answer in time constraint up to 10 minutes. In these questions the ability to understand structured sets of relationships, deduction of new information from sets of relationships, analysis and evaluation of the arguments, identification of central issues and hypotheses, and identification of plausible causal explanations is expected. Questions in the analytical section usually measure reasoning skills developed in virtually all fields of the study course.

From the types of questions we have mentioned above, the system for eTesting has implemented the following types of questions:

- multichoice questions;
- short answer questions;
- essay answer questions.

The main characteristics of multichoice questions is that students are choosing one or more answers from those which are offered as possible true answers. The number of offered answers is specified by the course administrator in the process of question design. It is recommended that every question should have more than 2 possible answers (in order to protect easy elimination) and less than 8-9 (the question can be displayed on one screen and there will be no need to scroll the page in order to see all possible answers). Every question must have at least one true and one false answer. While designing the possible answers it should be taken in consideration that the false answers should not be easily recognized by the students. Easy recognition of the false question can lead the students to the true answer using elimination. Questions can reference graphs, tables and diagrams. The student is making the choice using its mouse, clicking on the buttons in front of every possible answer.

Short answer questions are type of questions where students are required to enter short numerical value or short text. These questions can be used to fill the text of some question, or to enter result of some mathematical assignment.

Because of the complexity of the answer analysis, these questions do not have ability to assess the last three levels of Blooms taxonomy. One of the design considerations which must be taken while designing this type of questions is that the answer should be short and concrete.

When students are dealing with essay questions they are supposed to provide their answer in one or more sentences. According to Race [41], this type of questions “gives bigger opportunity for students to express their individualism”. The system we have designed there is no possibility to evaluate those answers automatically at the moment, because of the complex statistical and lexical analysis needed. The evaluation of these answers is done by the course administrator when the test is completely finished. Because the evaluation of these questions is manual, more time is required to provide feedback to students, which makes this type of questions not recommendable for tests which goal is to provide fast feedback.

The use of essay questions lowers the test objectivity comparing to the questions with fixed answer, because the person which is evaluating the answers has influence on the final result. “They are form of assessment in which the danger of subjective marking is at highest level” [41]. Another negativeness is the influence which has the ability to use the computer keyboard on the final results, where those students which type faster are favorite. The advantages of this type of questions are: the possibility to check student skills to compose and answer, the lower time to prepare the questions, and the possibility to check the higher levels of Bloom taxonomy.

#### *Test creation algorithm*

Test creation algorithm is closely connected to the chosen method for test delivery. The idea for creating different tests for every student, forced us to apply the model for dynamic test creation. With that idea every student will get different test, with same weight like all other student. These dynamically created tests will have fixed number of questions because this was first time system for automated assessment to be applied at the University. In order to provide less painful change in the way of taking the assessment and to lower the difficulties in its adaptation we have decided that fixed number of question is better solution than dynamic one. The same reason was forcing us to use dynamic test creation model instead of model for adaptive testing because of the easiness and the transparency non adaptable test have. The applied model gives opportunity for students to list the questions one by one, and answer only those which answer they know.

The strategy for test generation is defined from course administrator when he schedule the assessment. When setting the strategy course tree structure is used.

The administrator is marking the learning objects from which questions will be selected, specifying the number of questions taken from every learning object. This way the course administrator will have control over the curricula for which student knowledge will be assessed. Because every learning object has questions with same weight, the tests which will be generated will have same weight too, but the students will get different tests from those learning objects selected by the course administrator. The system has a feature with which already made strategies are saved and can be used in the future.

#### *Display systems*

The use of different computers and operating systems while using the system, forces adaptable user interface which will be able to adapt to different computer configurations. The server layer of the system works on Windows operating system installed on Pentium 3 with 128 ram memory, and uses SQL database. The questions are displayed one by one using screen resolution 800x600. Using this strategy the need for scrolling the screen in order to read question content is minimized. When questions are displayed, pictures, video and audio materials can be used.

#### *Marking and results reporting*

The system for eTesting implements evaluation of the entered answers at the end of the test, at the moment when the person which knowledge is tested specifies that the entered answers are definite. Because the most questions used in the system are fixed response questions, they are easy to evaluate when final results are displayed. The system displays the final results with an option to see the right answers compared to those entered by the user.

In order to eliminate guessing, we have implemented negative marking. The way we have implemented it is like follows. If number of possible answers is  $n$ , the number of true answers  $t$ , and the number of points  $p$ , then if the student chooses all true answers he should have  $p$  points, and if he chooses part of true answers then he is supposed to get part of the points. The number of points which are given for selecting any true answer are  $p/t$ .

$p/t$  - number of points which are given for selection of any true answer

$p/(n-t)$  - number of points which are taken for selection of any false answer

Any question should take maximum  $p$  points from students score if the one select only those questions which are false. That takes us to number of points which are taken from student for selecting any of the false answers offered for the question. Any false answer will take  $-p/(n-t)$  point from student score. The student which will check all possible answers will have 0 points to that question.

The final score of every test can be displayed on two ways, either using points or percents. The system we have created displays the final result as percent of possible points, at the same time showing the points which student got and comparing those

to the maximum points on the test.

## 5 Results

Large number of authors points out that the extensive growth of the information technology scientifically increases the use of computer-based testing as effective tool for knowledge assessment of a large group of students. At the same time, there are many studies which are focused on the implementation of the computer-based testing, and the results of that implementation. Usually the studies analyze use of existing software packages in high schools [42] and universities [43]. According to the results from these researches, the use of computers and web technologies in the assessment process, compared to the traditional way of assessment, are improving student achievements [43]. As Bocij and Greasley conclude [43], with using computer-based testing the effectiveness is increased and the time for test completion is decreased, allowing more time for thinking and solving the problems.

This paper focuses on a study to assess the effectiveness of the implementation of the system for computer-based testing at the PMF Institute of Informatics. The global goal is to identify the influence which the use of a system like this have on students and teachers, and the main goal was to see if the use of computer-based testing is more effective and more objective than traditional one.

Since we are discussing the effectiveness of the system which is implemented, and its influence in the assessment process, we must define what we will understand under the term effectiveness, because this term is defined differently depending on different points of view. According to Oxford, Cambridge and RSA Examinations, one of three main commissions for assessment in United Kingdom, few criteria exist which every software for computer-based assessment should achieve in order to be assumed as effective. The main criteria's are: cost reduction and quality improvement of the assessment process.

The quality improvement can be analyzed from two aspects, from teachers' point of view and from students' point of view.

The system for electronic testing eTest is implemented at the PMF Institute of Informatics in 2001 and has been used for summative and formative assessments usually when student number exceeds 100, often going up to 150. In order to assess more students at the same time, the summative assessments are carried out in 3 computer labs at the same time with exact time schedule. During the assessment process, besides the security functionalities implemented in the software, other security steps are taken necessary to avoid the traditional forms of cheating. The summative assessments are supervised by Institute staff.

The analyze of the results from its implementation and its use will be held from



the experience we have using the system at the Institute of Informatics during these 6 semesters of its use. In that period, the system was used for summative assessments in 12 courses, 327 assessments were held, and 5291 tests were generated during these assessments.

#### *Cost reduction*

One of the main results that should be achieved with the automation of the assessment process should be cost reduction, as is usually case with all automation procedures. In some cases, cost reduction is not possible for short period, especially when in the automation process software development and maintenance is needed, and training of those which will use the software must be obtained.

With the implementation of the system for electronic testing eTest, existing way of receiving assignments and questions on paper is changed with possibility for electronic delivery of questions and answers. With this, the use of paper is significantly reduced, and all printing resources for assessment preparation are eliminated. But, the use of systems like this one may cause lots of technical problems, which brings on the need for constant technical personnel, which is increasing the costs for its implementation.

The technical infrastructure can also be a problem on places where such infrastructure does not exist. The need of such infrastructure may require big investment by the customer.

Although it is very hard, maybe impossible to predict the time needed for return of the investment, because of the long term use of the systems for computer-based assessment, it is expected that the investment will be returned with time.

At the time when we started the development of the software, at the Institute of Informatics, there was already good existing technical infrastructure, which means that there was no need for additional hardware investment in order to implement the system. The costs for software development were also not very high, because its development by staff from the Institute. These facts, and the expectations that the teachers and students will benefit with this solution, was good motivation for its implementation at the Institute.

#### *Teacher perspective*

The main goal for development of a system for electronic testing was to offer faster and more objective assessment of student knowledge, to deliver faster and personalized feedback about students' achievements, which with the existing number of students at the Universities become almost impossible. The use of a system like this, should take some parts of the teachers job, giving them more time for other things.

The use of the system at the Institute was evaluated during its use, mainly through informal interviews with the faculty members which were using the system for creation of tests. From these interviews we concluded that the time teachers

spend preparing the assessments, was are not significantly reduced at the beginning of their use of the system. At the very beginning of every course, a lot of time was needed to create the question bank for that course, in order to achieve consistency and security of the assessment. The process of question bank creation usually took one semester for every course. In the next semesters only short attention was needed for new question implementation and maintenance of the existing ones. The number of questions for some of the courses at the moment is more than 1700, and the whole question bank at the moment has 9735 questions. As Bull points out [44], question design is one of the most time consuming actions which needs to be taken at the beginning of use of any system for electronic testing. Further she concludes that from time point of view, the achievements can be seen with longer use of these systems. From our experience, we have concluded that there was time reduction after the complete course organization. When any course is completely organized, the assessment scheduling and evaluation takes minimal time. Also, because of the fast, detailed and personalized feedback which this system gives to the students, more time reduction was achieved.

#### *Student perspective*

In order to check the student opinion about the effectiveness of the system for electronic testing, a research about their opinion is help with the use of questionnaire, a method widely used in the education for doing research. We used this method because with its use, we were able to gather information about student opinion. The most of the questions asked we-re taken from already held questionnaires, according to Miller and Engemman [45], and Bocij and Greasey [43]. The most of the questions are Likert type with 5 values (1-5). The group to which this questionnaire was intended was those students which knowledge was assessed with this system before.

The results from the 235 student questionnaires were used to evaluate student attitudes and perceptions toward automa-tion of the assessment process with the use of computers in it. The students were evaluating the easy of use, marking objectivity, opportunity to achieve more knowledge and the advan-tages of this form compared to the traditional forms of asses-ment. Results show that the student opinion was positive, and that they have accepted the use of the system, and they recommend its future use in other courses.

## **6 Conclusions**

The area of global testing is developing and further supported by the International Test Commission [46]. Computer-based testing, including PC-based, web-based and computer adaptive testing, has seen a continuous increase since 1980s [47].

The Association of Test Publishers has developed Computer-based Testing standards [48] [49]. A good survey of mega trends towards personnel testing, whether used for job-hunting, interviews; scoring etc. is given in [50].

In this paper we have analyzed the basic concepts of the systems for automated assessment. We have reviewed the possible implementations of these systems and mentioned their advantages and disadvantages. The results from these analyses were used in the process of definition and development of new system for automated assessment - eTest. eTest is Web based system for electronic testing, which consists of few modules. The main concept behind this system is the question bank, from which questions are drawn when tests are generated. The system have implemented algorithm for test creation according to which different tests are created dynamically with same weight according to some previously predefined strategy. Generated tests have fixed length and give opportunity for browsing the questions before answering them.

Because this was the first time students from the Institute of Informatics to work and be assessed on system like this, we have concluded that it will be more accepted if the way of assessment have some similarities with the old system we have used before on paper. That is why we did not implement a module for adaptive testing in this phase. Implementing a module for adaptive testing at this stage can result with negative reactions from students, because they will not be able to see that number of question on the test and because of the additional explanation which will be needed about the way the system selects the question from the question bank. The use of module for adaptive testing requires calibrated question bank which we did not have at the beginning of this project. Using this platform we have made some analysis about the influence which technology have in the process of assessing someone's knowledge.

The implementation of the system for electronic testing on the Institute of Informatics, confirmed already existing constations of the researchers in this field [44], which say that the use of technology in the education is very dependent of the organization in which it is implemented, the logistics that organization will provide, and the cultural changes which it will arise it that organization. Because on the Institute we do not have synchronized systems for management of the students, we have faced with serious problems in many occasions which we have solved successfully. The positive experience in implementation of the system was the existence of complete and sophisticate computer equipment at the Institute which is one of the main prerequisites for using systems like this. If infrastructure like this does not exist, then the investment which is needed can demotivate the institutions for implementing a system for automated assessment.

The system for electronic testing eTest is used for assessment of student knowledge on the Faculty of natural sciences and mathematics in period of 6 semesters.

During this period 327 assessments have been made on which 5291 student took part. The study on using computer-based assessment software as a tool for knowledge assessment, was based on student experience evaluated using a questionnaire, and on informal inter-views with the faculty members which are using the implemented system.

The results from the informal interviews are showing that a system like the implemented one, is giving more opportunities to the teachers, slightly reducing their time spent on assessment preparation and evaluation of the student answers. The results show that the time reduction is shown after using the system for some time, not immediately. Using the system the problem we were facing with when assessing large number of students very often was overcome. We have even shortened the time for providing the student with feedback about the assessment by implementing the module for automated evaluation of their answers. The results are showing detailed information about the true answers and analyze of the student answers compared to the true ones. The student interest was obvious in analyzing their answers and comparison with the true ones. Another achievement of this system was the possibility for statistical analyses of the gathered responses from the students. Using the modules for statistical analyses, it is very easy to identify the weak questions and the parts of the curricula which are not very clear to students. This information can be used in the process of teaching.

The results from the 235 student questionnaires were used to evaluate student attitudes and perceptions toward automation of the assessment process with the use of computers in it. The students were evaluating the easy of use, marking objectivity, opportunity to achieve more knowledge and the advantages of this form compared to the traditional forms of assessment. Results show that the student opinion was positive, and that they have accepted the use of the system, and they recommend its future use in other courses. Especially they liked the fast feedback on their answers which filled the lack of conversation between the students and the teachers. The results have shown that those student which used the module for learning thought assessment and the module for self assessment achieve better results than those which did not use those modules.

The use of the system, lowers but have not completely removed the classical ways of cheating, at the same time opening some new ways of cheating mostly characterized by the use of technology. We were witnesses of many tries from the students to cheat the system in order to achieve better grade, but most of those attempts were discovered on time.

The directions for the future work can be divided in few segments: improvements in the concepts on which the system is based and improvement in its implementation. We will try to provide compatibility with the other existing systems for automated assessment which may require restructuring of the question bank, but it

will allow importing and exporting questions from and to other systems. Besides the existing types of question we have implemented, we will try to implement other types of questions and make an attempt to create a module for automated evaluation of free response questions. The use of adaptive testing is one of the test creation algorithms which will be implemented in the future which will improve the process of assessing someone's knowledge.

## References

- [1] S. Davis and J. Botkin, *The Monster Under The Bed*. Simon & Schuster., 1994.
- [2] A. Toffler, *Future Shock*. London: Pan Books, 1972.
- [3] C. I. E. Copyright. Compton's NewMedia, 1995.
- [4] D. P. Diaz and K. F. Bontenbal, "Pedagogy-based technology training," in *Teaching and Learning in a Network World*, P. Hoffman and D. Lemke, Eds. Amsterdam, Netherlands: IOS Press., 2000, pp. 50–54. [Online]. Available: [http://home.earthlink.net/davidpdiaz/LTS/pdf\\_docs/teched2K.pdf](http://home.earthlink.net/davidpdiaz/LTS/pdf_docs/teched2K.pdf)
- [5] R. W. Tyler, "The five most significant curriculum events in the twentieth century," *Educational Leadership*, vol. 44, no. 4, pp. 36–38, Dec./Jan. 1986-87.
- [6] A. L. Brown, D. Ash, M. Rutherford, K. Nakagawa, A. Gordon, and J. C. Campione, "Distributed expertise in the classroom," in *Distributed cognitions: Psychological and educational considerations*, G. Salomon, Ed. New York: Cambridge University Press, 1993, p. 188228.
- [7] P. Cobb, "Where is mind? constructivist and sociocultural perspectives on mathematical development," *Educational Researcher*, vol. 23, no. 7, p. 1320, 1994.
- [8] M. Dougiamas, *A journey into constructivism*, 1998. [Online]. Available: <http://dougimas.com/writing/constructivism.html>.
- [9] J. Bruner, *Going Beyond the Information Given*. New York: Norton, 1973.
- [10] B. R. Snyder, *The Hidden Curriculum*. Cambridge: MIT Press, 1970.
- [11] S. Brown and P. Knight, *Assessing Learners in Higher Education*. London: Kogan Page, 1994.
- [12] K.P. Cross and M.H. Steadman, *Classroom research: implementing the scholarship of teaching*. San Francisco: Jossey-Bass, 1996.
- [13] University of new england assessment policy. [Online]. Available: [http://www.une.edu.au/offsect/assessment\\_policy.htm](http://www.une.edu.au/offsect/assessment_policy.htm)
- [14] R. Waters and M. McCracken, "Assessment and evaluation in problem-based learning," presented at the Frontiers in Education Conference, Pittsburgh, USA, nov 5-8 1997. [Online]. Available: <http://fie.engrng.pitt.edu/fie97/papers/1454.pdf>
- [15] L. B. Darrell, "The impact of computer-based testing on student attitudes and behavior," *The Technology Source*. [Online]. Available: <http://ts.mivu.org/default.asp?show=article&id=1013>

- [16] D. Natal, "On-line assessment: What, why, how." presented at the Technology Education Conference, Santa Clara, California, May 6 1998, pp. 1–23.
- [17] *National Research Council. High Stakes: Testing for Tracking, Promotion, and Graduation.* Washington, DC: National Academy Press, 1999.
- [18] R. Grabinger, "Environments for active learning," in *Handbook of Research for Educational Communications and Technology*, D. Jonassen, Ed., pp. 665–692.
- [19] R. E. Bennett, "Using new technology to improve assessment," *Educational Measurement: Issues and Practice*, vol. 18, no. 3, pp. 5–12, 1999.
- [20] J. Bull, "Update on the national tlt3 project: The implementation and evaluation of computer-assisted assessment, keynote," in *Proceedings of the 3rd Annual Computer Assessment Conference*, Loughborough, UK, June 16–17, 1999. [Online]. Available: <http://www.lboro.ac.uk/service/lt/flicaa/conf99/pdf/keynote.pdf>
- [21] P. Summons, J. Coldwell, C. Bruff, and F. Henskens, "Automated assessment and marking of spreadsheet concepts," presented at the Proc. of the Second Australasian Conference on Computer Science Education, Sydney, Australia, July 3–5, 1996, pp. 178–184. [Online]. Available: <http://portal.acm.org/citation.cfm?id=299385&jmp=cit&dl=guide&dl=acm&cfid=11111111&cftoken=222222>
- [22] (2001) School technology and readinesskey building blocks for student achievement in the 21st century: Assessment, alignment, accountability, access, analysis, 2001. CEO Forum on Education and Technology. [Online]. Available: <http://www.ceoforum.org/downloads/report4.pdf>
- [23] M. Russell and W. Haney, "Bridging the gap between testing and technology in schools," *Education Policy Analysis Archives*, 2000. [Online]. Available: <http://epaa.asu.edu/epaa/v8n9.html>
- [24] R. E. Bennett, "How the internet will help large-scale assessment reinvent itself," *Journals in Education*, vol. 9, no. 5, Feb. 14, 2001. [Online]. Available: <http://epaa.asu.edu/epaa/v9n5.html>
- [25] S. Behrens. Online assignment management: an evolutionary tale. Central Queensland University. Rockhampton, Australia. [Online]. Available: <http://dlib.computer.org/conferen/hicss/1874/pdf/187450156c.pdf>
- [26] Caa centre website. [Online]. Available: <http://caacentre.ac.uk/>
- [27] J. Olsen, "Guidelines for computer-based testing," *Association of Test Publishers (ATP)*. [Online]. Available: <http://www.isoc.org/oti/articles/0500/olsen.html>
- [28] M. Freewood and J. Hanson. Computer assisted assessment (caa) literature review. School of Computing and Management Sciences in October 2000. [Online]. Available: <http://www.shu.ac.uk/services/lti/resources/caapracticalguidance/>
- [29] B. Landon. Online educational delivery applications: A web tool for comparative analysis. [Online]. Available: <http://www.c2t2.ca/landonline/index.html>
- [30] (2001, oct) Ims global learning consortium. ims question and test interoperability specification. [Online]. Available: <http://www.imsglobal.org/question/index.cfm>

- [31] C. F. "Web hosted assessment-possibilities and policy," in *Proceedings of the 6th annual Conference on the Teaching of Computing/3rd Annual ITiCSE Conference on Changing the Delivery of Computer Science Education*, 1998, p. 5558. [Online]. Available: [http://portal.acm.org/citation.cfm?id=283020& jmp=references& dl=guide&dl= acm& cfid=11111111& cftoken=2222222](http://portal.acm.org/citation.cfm?id=283020&jmp=references&dl=guide&dl=acm&cfid=11111111&cftoken=2222222)
- [32] T. Patelis. An overview of computer-based testing. The college board. RN-09. [Online]. Available: <http://www.collegeboard.com/research/html/rn09.pdf>
- [33] B. Brent, L. Lou, and J. Altamese, "Effects of screen size, screen resolution, and display rate on computer-based test performance," presented at the Annual Meeting of the National Council on Measurement in Education. New Orleans: Educational Testing Service, Princeton, New Jersey, april 2002. [Online]. Available: <http://www.ets.org/research/dload/AERA2002-bri2.pdf>
- [34] B. G. Dodd and S. J. Fitzpatrick, "Alternatives for scoring computer-based tests," invited paper, presented at the IEducational Testing Service's Computer Based Testing Colloquium: Building the Foundation for Future Assessments Colloquium, Philadelphia.
- [35] M. Pettigrew. Random guessing on multiple choice tests. [Online]. Available: <http://www.shu.ac.uk/services/lti/people/mp/mcq/>
- [36] S. Watts, "Integrated computer-assisted assessment," *Computer-Assisted Assessment Evaluation and Survey Service (CAAESS)*. [Online]. Available: <http://www.lboro.ac.uk/service/ltd/flicaa/conf2001/pdfs/j1.pdf>
- [37] P. Davies, "There's no confidence in multiple-choice testing," School of Computing. [Online]. Available: [http://www.lboro.ac.uk/service/ltd/flicaa/conf2002/pdfs/davies\\_p1.pdf](http://www.lboro.ac.uk/service/ltd/flicaa/conf2002/pdfs/davies_p1.pdf)
- [38] J. McKendree, "Negative marking and gender bias," CAAC, 2002. [Online]. Available: [http://caacentre.lboro.ac.uk/resources/faqs/negative\\_marking1.shtml](http://caacentre.lboro.ac.uk/resources/faqs/negative_marking1.shtml)
- [39] M. Gusev and G. Armenski, "New frontiers for testing in e-learning," in *Proceedings of SSGRR, Conference for Advances in Infrastructure for Electronic Business, Science and Education*, Roma, 2001.
- [40] Description of the general test. Graduate Record Examination. [Online]. Available: [www.gre.com/cbtttest.html](http://www.gre.com/cbtttest.html)
- [41] P. Race, "Art of assessing," *The New Academic*, vol. 4,3 & 5,1, 1995. [Online]. Available: [http://greguns2.gre.ac.uk/pcet/programmes/celtt/celtres.nsf/0/51c5bc41530d44c480256c7e004493a1/\\$file/the+art+of+as sessment.pdf](http://greguns2.gre.ac.uk/pcet/programmes/celtt/celtres.nsf/0/51c5bc41530d44c480256c7e004493a1/$file/the+art+of+as sessment.pdf)
- [42] M. Russell, "Testing on computers: A follow-up study comparing performance on computer and on paper," *Education Policy Analysis Archives*, vol. 7, no. 20, 1999. [Online]. Available: <http://www.bc.edu/research/intasc/PDF/TestCompFollowUp.pdf>
- [43] P. Bocij and A. Greasley, "Can computer-based testing achieve quality and efficiency in assessment?" *International Journal of Educational Technologies*, vol. 1, no. 1, 2000. [Online]. Available: <http://www.ao.uiuc.edu/ijet/v1n1/bocij/>

- [44] J. Bull, "Computer based assessment: some issues for consideration," in *TLTP Project ALTER*. [Online]. Available: <http://www.ilt.ac.uk/public/cti/ActiveLearning/issue1/bull/>
- [45] H. Miller and K. Engemann, "The effectiveness of web-based testing in higher education: A case study," in *Proceedings of the Decision Sciences Institute 31st Annual Meeting*, Orlando, Florida, November 18-21, p. 2000. [Online]. Available: [http://www.sbaer.uca.edu/research/2000/dsi/pdffiles/papers/volume1/pt3/v1\\_104.pdf](http://www.sbaer.uca.edu/research/2000/dsi/pdffiles/papers/volume1/pt3/v1_104.pdf)
- [46] (2000) International test commission, international guidelines for test use. Netherlands. [Online]. Available: <http://www.zpid.de/en/products/pub/testguide.php>
- [47] H. Lyman, *Test Scores and What They Mean*, 6th ed. Boston: Allyn & Bacon, 1998.
- [48] *Association of Test Publishers: "Guidelines for Computer-Based Testing"*. Washington, D.C.: Association of Test Publishers, 2000.
- [49] W. Haris, "Best practices in testing technology: Proposed computer-based testing guidelines," *Journal of e-Commerce and Psychology*, vol. 1, no. 2, pp. 23–35, 2000.
- [50] J.W. Jones and K.D. Higgins, "Mega trends in personnel testing: A practitioners perspective," *Association of Test Publishers*, jan 2001. [Online]. Available: <http://www.testpublishers.org/journal03.htm>