Chapter 17

ONLINE EVALUATION AT HIGHER AND SECONDARY EDUCATION

A teaching and learning experience

Joglar, N., Risco, J.L., Díaz, A., Colmenar, J.M.
Ingeniería Técnica en Informática de Sistemas

Centro de Estudios Superiores Felipe II, C/ Capitan, 39, 28300 Madrid, Spain

Abstract: This chapter describes the experience of developing and testing a tool for online evaluation in academia. The tool was firstly built as a result of a collaboration project among several subjects in the Computer Science Department at the college C.E.S. Felipe II (Aranjuez, Spain). A later beta-version of the application was tested in the evaluation of students from five different courses like *Music* at the middle school level or *Calculus* at college level. We include here a brief summary with the results of all tests. The absence of some features detected in those tests and the complexity of the tool, which was coded by students, guided the people in charge of the research team to decide the development of a new tool starting from scratch. The features of this new tool, called *iTest*, are also shown in this chapter, detailing the differences it presents in relation to the previous tools.

Key words: Online Evaluation, New Technologies in the Classroom, Mathematics Education, High School Education, Teacher Education.

1. INTRODUCTION

The instructional films of the 1940's were expected to radically change the educational delivery system, as were instructional radio and television. While each one of these technological innovations had some impact on educational programs, they did little to change the fundamental nature of education. The Internet and computer technology, as the next generation of

B. Fernández-Manjón et al. (eds.), Computers and Education: E-learning, From Theory to Practice, 193–204.

technological innovation to impact distance education, appears to have the power to alter the education landscape significantly (Aragon and Johnson, 2002).

The Internet offers the possibility of providing a stimulating environment to engage students in meaningful learning through reflection, application and interaction (Macdonald et al., 2001). In addition, web-based quizzes provide teachers and students with several advantages as the technology for self-assessment (Chapman, 2004). For these and other reasons, the number of web-based or online courses has increased rapidly over the past few years to become a significant constituent of educational pedagogy worldwide. Online learning or e-Learning is stimulated primarily by using new technologies (Suanpang et al., 2003).

Conformance to e-Learning standards, such as IMS (Instructional Management System) (IMS, 2006) and SCORM (Sharable Content Object Reference Model) (SCORM, 2006), is increasingly seen as the way forward in the development and deployment of e-Learning frameworks. The standards use XML technology for the specification and are released by the IMS consortium. One of these standards is Question and Test Interoperability (QTI), developed for constructing and exchanging tests and assessment information.

Our research team in the Computer Science Department of the higher education institution called CES Felipe II in Madrid, has always been conscious of the importance of the integration of e-Learning technologies. In this direction, we are developing a tool called *iTest*, which is a subset of a *Virtual Learning Environment* (VLE) or *Learning Management System* (LMS). In this chapter we present iTest giving all the details from its motivation to its technical features. iTest is an online management software that allows instructors to configure exams and students to take them. The tool was started with the aim of improving the current strategies to implement mathematical formulas for online exams on the web. The scope of iTest has been enlarged in order to store multimedia contents and to allow the design of *Music* course tests and animated tests using Flash. Our idea is also to include the feature of importing and exporting from IMS QTI 2.1. iTest is aimed to be part of a research project between our Institution and the Education Department of Aranjuez City Council. This project includes developing courses to instruct secondary school teachers on the usage of this tool next academic years.

The rest of the chapter is organized as follows: firstly, we describe the motivation of our work, then we give all the details of the development of the tool; after that, we present the results obtained using the tool in higher and secondary education. Finally, we give the conclusions of this work.

2. MOTIVATION

With the availability of course management software such as Blackboard (Blackboard, 2006), WebCT (WebCT, 2006) and many others, most of educators are converting conventional tests into web-based tests. As an example, the state of Virginia, USA, required all public schools to have the capability of online testing. Virginia students took over 226,000 online tests in 2004. The state of Indiana (USA) developed online end-of-course assessment in *Algebra I* and *English* for 11th grade students. About 91,000 students took the tests, and their schools received the results within 48 hours. Teachers appreciated the speedy feedback and the absence of test booklets (Rodriguez, 2005).

In Spain the situation is not the same. In spite of the rapid growth in the accessibility of new technologies, for example, the program *one computer per two students* in the Spanish region of Extremadura, the new facilities for online evaluation are not being used (Guijarro, 2005).

Commercial systems are usually very expensive to be implemented on a large scale and development costs are very high. Since they all use different formats, test libraries cannot be easily integrated or accessed hence reducing reuse and portability. In addition, teachers are often discouraged because they do not have time to learn how to use platforms with so many features.

In view of all these problems, one of the goals of our research team is to develop a new non-commercial tool, called iTest, which is easy to use and does not require any commercial software to run. Besides, formation as well as continuous assistance for teachers using the tool will be provided by our team members.

Through an innovative application and combination of open Web technologies such as MathML, Java active pages (JSP), webservices (J2EE) and database access (MySQL), iTest aims to create a website for online exams where contents and support services are interoperable. Contents and support services can also be created, deployed and maintained on a distributed basis. The iTest website is organized to systematically supplement complete courses curricula in any given school.

The idea of developing this project was suggested by the Mathematics lecturers at our Department. There was a strong need for a new teaching and learning methodology using new technologies to increase the motivation of the students. Different versions of a preliminary tool were developed in collaboration with some of our students during the academic years 2003-05 (Díaz et al., 2005; Joglar at al., 2006). iTest not only allows them to self-test their knowledge in a specific area at a given point, but it also increases the motivation for the students to study abstract and difficult concepts. This idea is clearly extensive to all levels of education.

2.1 Higher Education

Based on our experience, Computer Science students seemed to be highly interested in using this kind of tool, in particular during the first semester of their first year at the university, since they have not had the opportunity to work at a university computer laboratory yet. Besides, these students in particular are not scared to turn on a computer. In fact, they are really eager to work with one even to learn mathematics.

We find these tools essential to help us motivate our students to learn cold and abstract topics such as *Calculus*, *Discrete Mathematics*, *Linear Algebra* and *Statistics*. It is important to remark here that iTest lets us not only to visualize graphs of functions and to reproduce sound files, but it also allows us to properly display mathematical formulas without using screenshots. This issue is a real plus of the tool.

2.2 Secondary Education

Another goal of the project is to modernize the educational process in the secondary education system where the knowledge growth is rapid, and it demands constant curriculum redefinition and implementation of new methods for efficient learning.

The original idea to implement e-Learning in order to modernize the educational process, is growing as the new hardware and software tools are now available at school and it is expected that the teachers are motivated to use new technologies to improve their teaching. It has also been expected that the educational process can be significantly improved by creating new educational resources with visually appealing multimedia contents that include interactive elements and up-to-date reusable information, as in (Smothers et al., 2004).

3. DESCRIBING THE TOOL

In this section we explain the history of the development of the tool, its different features, the exam configuration management and some other technical details.

3.1 Development History of the Tool

In the next paragraphs we would like to carefully specify all the development stages of the tool and the different versions obtained.

During the 2003-2004 academic year, an online application that allows random generation and automatic grading of exams was developed through an educative experience between several subjects at our Department (Díaz et al., 2005). Some students had the opportunity to develop a medium size project instead of small exercises for each class working together and closely supervised by a group of professors from our Department. We think this is a very interesting aspect of our tool; it has been developed by students, and it is conceived to be used by the students themselves. In order to complete this first stage, a group of students voluntarily developed a first simplified operative version. They named the first version *ExaNet* (Díaz et al., 2005).

In May 2005 a first evaluation of this initial version is performed in the *Statistics* course at our Department. After this first evaluation, some problems were detected: low usability in the interface of the student taking an exam, administrator interface too simplified and mathematical formulas not supported.

The tool was partially improved by a second group of students again assisted by professors from our Department during the first semester of the academic year 2005-2006. Students were granted some free configuration credits for their work (Joglar et al., 2006). This second version of the tool will be referenced along the article as the beta-version of the tool.

In January 2006 and May 2006, tests were performed in the *Calculus* and *Statistics* courses, respectively. Exams at secondary school level are also carried out (one in an *Algebra I* course and another one at a *Music* course). The results obtained are very encouraging. Students are satisfied with the new evaluation system but some usability problems are detected in the administrator´s interface. The teachers using the tool could not work independently. They had to prepare their questions and answers in a separate document, and then the general administrator user was in charge of transferring all the data to the database. This general administrator was also the only user able to register new students, to configure exams and to access the statistics of the results.

After our experience with the beta-version over the past two academic years, we are nowadays developing a new application with all the features that have been detected as necessary for an optimum functionality (see next subsection). This new tool is being developed from scratch and it has been named *iTest*.

The development of iTest has been integrated in a project with the Aranjuez City Council in order to systematically use the tool at all local high schools and middle schools. We have also obtained financial support from the *Universidad Complutense de Madrid* through their program for S*pecial Actions in Research*.

In the next subsections we will present the features of iTest emphasizing the differences with the beta-version.

3.2 Features of iTest

Our tool considers three kinds of users: instructors, students and a general administrator user.

Instructors have to structure the subjects they teach in different didactic units. They may add, edit or delete any question or answer for each didactic unit within each subject. A level of difficulty can be assigned to each question. They may also configure any number of exams and visualize different types of statistics of the exams performed by the students.

Students may choose to start an exam from the list of courses that have any exam configured for the current date within the set of subjects the student is currently enrolled in. Once the student selects the subject, the tool randomly generates the exam for the chosen subject by using the configuration held at the database. If the assigned time length for the exam expires, the tool automatically finishes the exam and alerts the user with a message. The student is also able to finish the exam at any time. In both cases, the tool grades the exam and it automatically shows the user's grade on the screen right after completion. The tool saves the information corresponding to each generated exam to allow the student to revise his exam during the correction period, which starts the day after the exam is not available. The students may check different aspects about the exams they did: generated questions, given answers, correct answers, grade per question and final grade.

The general administrator user is able to add, edit or delete information stored at the database. Currently, we use the administrator features to add students, instructors and subjects to the database and to include the relationships between them.

The main difference between the beta-version and iTest is that the beta-version did not have instructor users. As we have already mentioned in the previous subsection, this is a crucial aspect of the tool in order for the teachers to be able to work almost independently from the general administrator.

3.3 Exam configuration and statistics

The instructors can configure more than one exam for the same subject during a course period. The configuration of each exam has the following configurable parameters: time to solve the exam in minutes, points per right answer, penalization for each wrong answer, dates in which the exam will be available for the students and information about the didactic units that are

covered by the exam. The configuration for a didactic unit consists of the next parameters: number of questions, number of answers per question and maximum and minimum difficulty of the questions. In this way it is possible to generate an exam that contains questions only from the chosen didactic units and within the selected difficulty level.

The beta-version did not have the possibility to decide which didactic units were covered by a given exam. However, the tool selected a balanced number of questions per didactic unit using all the didactic units included in the particular subject. This version did not have either the possibility of assigning a level of difficulty per question.

Given that the tool randomly generates each exam, it is very difficult that two students get the exact same exam at the same time. In spite of this fact, when using the tool via Internet without any supervision, it is difficult to control all the factors involved to avoid cheating. Thinking in a non-supervised use of this tool, the final version will register the IP address, time of starting the test and time of answering a given question.

The instructor user can also fix the specific grades assigned to each question at a given exam. When a question has more than one correct answer, the exam might be set up to only penalize the student with wrong answer if none of the correct answers is selected. If a student does not select all the correct answers for such a question, the penalization is proportional to the number of correct answers selected compared to the number of correct answers. For example, if a student selects one correct answer in a question with two correct answers from an exam where 1 point is given for a full correct answer and -0.3 per wrong answer, the final the penalization is -0.15 (-0.3/2). It might also be interesting to study giving partial credit for wrong answers even if there is only one correct answer per question. For that, one would have to think of an ordering in the set of possible answers for a particular question from *totally wrong to totally correct* (Ashton et al., 2006).

Finally, the tool also provides statistics like averages, standard deviations, percentages of correct answers for a given question, etc. In particular, the tool computes the ratio of success for a given question, that is, number of students who got the right answer divided by the number of students who had that question on the exam. These ratios are important because they allow the instructor to detect misunderstandings by the students in a particular point of the course syllabus.

3.4 Technical details

All the tools described above were developed as Java applications, based on a MVC (model-view-controller) architecture. The database support has been developed according to the SQL standard.

Going a step further, the Java application was converted into a dynamic web application, using J2EE platform. Since it was developed according to MVC paradigm, only view and controller had to be redesigned using J2EE technologies: JSP and servlets. The web application was deployed into a Tomcat application server.

The most innovative aspect of the tool is the support for visual representation of mathematical formulas. This feature was implemented using MathML. Instructors are able to include, in questions and answers statements, mathematical formulas written in a LaTeX-like notation. Then, we reused a JavaScript code written by Peter Jipsen to transform this notation into MathML (AsciiMathML, 2006). Client browsers can then display the formulas if they have the appropriate plug-in (MathPlayer, 2006). Other resources, such as sound tracks or flash animations can also be included in questions and answers. For these cases the tool just stores the file and embeds it into the HTML presentation in order to be described by the adequate plug-in of the client browser.

4. TESTING THE BETA-VERSION OF THE TOOL IN HIGHER EDUCATION

The beta-version of the tool described above has been successfully tested with first and second year Computer Science students at our university since May 2005. All our students have taken the online tests in Math courses voluntarily. They also took the test in computer laboratories at our Institution supervised by their instructors, except for one test in the *Calculus* course. *Calculus* students took the test online at home without any kind of supervision in January 2006.

All students who took an online test filled in an evaluation form right after finishing their tests. All participants agreed that their experiences had been very interesting and motivating for them. The most appreciated feature of the tool was the fact that the tool gave them their grades immediately. Many of them suggested a methodic and continuous use of the software to help them self-test their knowledge of a given topic at any given time.

We remark that we are not in the position of making a formal statistical study of the usage of this tool yet because we have conducted only one basic test with each participating group of volunteered students. Once we check the well-functioning of the final version, we will be ready to design formal usages of the software.

In Table 1 we include significant data from the three tests we have conducted with our students from May 2005 until May 2006. Two of them

Table 1 Description of different tests at university level

Course	#questions; #answers; time	Didactic units	Participation	Grading	Passed
Statistics (May 05)	4; 5 per question; 15 min.	2	17% (18/105)	0,25;-0,1	100%
Calculus (Jan. 06)	5; 5 per question; 30 min.	3	60% (58/98)	0,2;-0,1	43%
Statistics (May 06)	5; 4 per question; 25 min.	3	42% (44/104)	0,2;-0,1	48%

were given at the end of the *Statistics* course during two academic years. *Statistics* is a mandatory 2nd year course covering 7.5 credit hours work (75 course hours) during the Spring semester. Students took these tests voluntarily at our computer labs supervised by their instructors. The grade obtained by each student in this exam was agreed to be add up to the grade in the final exam (grading scale of the final: from 0 to 10). On the other hand, *Calculus* students took their tests voluntarily at home without any supervision. *Calculus* is also a major course for our first year students with 6 credit hours work (60 course hours) during the Fall semester.

All these tests were design with only one correct answer per question. Partial credit was not given for wrong answers. Students were penalized for wrong answers to avoid random responses and students were not penalized for leaving a question blank. In the grading column of Table 1, we specify the number of points given per correct answer and the number of points subtracted per wrong answer. All questions of a given exam had the same level of difficulty.

5. USAGES OF THE TOOL IN SECONDARY EDUCATION

The beta-version of the tool has also been tested at secondary education level in collaboration with two secondary schools of our geographic area. Since one of the main functionalities of the tool includes the editor for math formulas, we decided to design at least one test for 11th grade *Algebra I* students from the *Instituto de Enseñanza Secundaria Fortuny*. All students involved in this particular test were science majors. They all were enthusiastic with the experience. Students appreciated especially the fact that the tool was developed by computer science students at a nearby college. They all affirmed that they would like to use this software in a continuous way to help them self-test their knowledge in a given subject at any time during the semester. This feedback gave us the idea to prepare an experiment to be completed during the academic year 2006-2007 with 12th grade students adapting questions from the national exams to access public universities in Spain.

On the other hand, given that it is also possible to include sound files in the data base of our system, we decided to run a test for a *Music* course. We design a simple test for 8th grade students with the help of the *Music* teachers of a local middle school (*Instituto de Enseñanza Secundaria Las Salinas*). Again, students were very satisfied with their performances in the exam, except for two of them who misunderstood the rules for validating their answers.

In these two experiments at local secondary schools, we decided not to penalize students for getting a wrong answer. To do so, we set up the tool so that it added 1 point per correct answer and 0 points per wrong or blank answer to the student grade.

The results of both tests are included in Table 2 following the same structure as Table 1:

Table 2 Description of different tests at secondary school

Course	#questions; #answers; time	Didactic units	Participation	Grading	Passed
Algebra I (Mar. 05)	8; 4 per question; 50 min.	4	90% (28/31)	1;0	82%
Music (May 06)	6; 4 per question; 20 min.	3	100% (24/24)	1;0	92%

It is worth mentioning that the passing rate was extremely high in both cases. Despite the lack of statistically significant data at this level, all the parties involved in the experiments agreed on the positive educative aspects of the usage of this tool as an assessment tool.

The teachers were a little concerned at the beginning with their lack of knowledge of the management of the tool. They all overcame this feeling with the continuous support offered by the members of our research team.

We deeply believe that it is crucial to engage students in critical thinking especially at the secondary education level nowadays. For that it is important that they get used to different evaluation systems as well as to the progressive integration of new technologies in the classroom.

There are a lot of improvements to make in this direction. Teachers at primary and secondary schools in our country feel often discouraged when it comes to the use of new technologies in the classroom. It is true that there are many tools available for them through the Spanish Ministry of Science and Education. However, it is very hard, some times even impossible, for them to use these applications without close supervision and permanent personalized assistance. Many tools are too wide and complicated to use. They offer functionalities that the teacher might not need and this fact makes its practical use in the classroom even harder for the instructor.

We believe that it is important to study how to build an effective system of preparation and professional development for general teachers concentrating on their continuous formation and technical support.

6. CONCLUSIONS AND FUTURE WORK

After all the preliminary study described above with the beta-version of the tool including the five tests that have been run, we are developing iTest, the new tool that implements all the extra features discussed in section 3.

We are also in the position of designing a formal study of an extensive use of the system to obtain statistically significant results.

In order to do so, we are establishing a cooperation agreement between our Institution and the Education Department of Aranjuez City Council. This agreement includes a project with the regional resource center (*Centro de Apoyo al Profesorado*) in Aranjuez to develop courses to instruct secondary school teachers on the use this tool this coming academic year 2006-2007. The courses will be taught by professors from our research team at our Computer Science Department. There will also be technicians working full time to be sure that the tool always works properly and to implement the new needs that could arise while the teachers are extensively using the software with their students. They will also provide continuous assistance to the teachers using the tool.

The flexibility of this type of systems is so big that its use at other educational levels (primary education, adult education, students with special needs, etc) could be easily given with the appropriate equipment and with the adaptation of the formation courses for the teachers to the precise educational level.

Signing this kind of institutional agreements will build affiliations and promote dialog among individuals from different statements to affect positive change in the academic community of the city of Aranjuez.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to all the students and instructors involved in all the tests mentioned in this article. The nature of our project allowed us to work effectively from the geographic distance with lots of people from different institutions without necessarily meeting in person.

We would especially like to thank the following people:

Former student advisors: D. J. Bodas, R. Sánchez and F. J. Soltero, instructors at the Computer Science Department, CES Felipe II.

Student development Groups: J. Barba, B. Blanco, D. Jiménez, F. Pérez, V. Sánchez, D. Sanz and J. Fernández.

Secondary education teachers: V. Díaz from IES Las Salinas, Seseña, Toledo, Spain and A. Corbato from IES Fortuny, Madrid, Spain.

This work has been partially supported by the program *Acciones Especiales AE-5/06-15073 del Vicerrectorado de Investigación de la Universidad Complutense de Madrid.*

REFERENCES

Aragon S., and Johnson S., 2002, An Instructional Strategy Framework for Online Learning Environments, in: *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* edited by G. Richards (Chesapeake, VA: AACE), pp. 529-536.

Macdonald C. J., Stodel E. J., Farres L. G., Breithaupt K., and Gabriel M. A., 2001, The demand-driven learning model: A framework for Web-based learning, Internet and Higher Education 4, 9-30.

Chapman N. S., 2004, Various ways of Using Online Testing in a Virtual Learning Environment, Brookes eJournal of Learning and Teaching, 1(1), 1-5.

Suanpang P.,Petocz P., and Kalceff W., (2003) Student Attitudes To Learning Business Statistics Online Vs Traditional Methods, in: HERDSA 2003 Learning for an Unknown Future (Christchurch, New Zealand).

IMS, 2006; http://www.imsglobal.org

SCORM, 2006; http://www.adlnet.gov/scorm/index.cfm

Blackboard, 2006; http://www.blackboard.com

WebCT, 2006; http://www.webct.com

Rodriguez N. C., 2005, State to try some CATS exams online, hopes for quicker results, in: The Courier-Journal.

Guijarro J. V., 2005, The global project on the information society in Extremadura: a free step towards future, in: Information Technology Interfaces, pp. 347- 351.

Díaz A., Colmenar J. M., Risco J. L., Joglar N., Sánchez R., Bodas D. J., and Soltero F. J., 2005, Experiencia educativa entre varias asignaturas, in: Actas de las XI Jornadas de Enseñanza Universitaria de Informática (Thomson).

Joglar N., Risco J. L., Sánchez R., Colmenar J. M., and Díaz A., 2006, Experiencia en el desarrollo y utilización de una herramienta de corrección automática de exámenes, in: Actas de las XII Jornadas de Enseñanza Universitaria de Informática.

Smothers V., Greene P., and Cohen C., 2004, Implementing an E-Learning Strategy, in: Strategic Reports from MedBiquitous Consortium.

Ashton H. S., Beevers C. E., Korabinski A. A., and Youngson M. A., 2006, Incorporating Partial Credit in Computer-Aided Assessment of Mathematics in Secondary Education, British Journal of Educational Technology 37(1), 93-119.

AsciiMathML, 2006; http://www1.chapman.edu/~jipsen/asciimath.html

MathPlayer, 2006; http://www.dessci.com/en/products/mathplayer