Online vs. Paper-Based Testing: A Comparison of Test Results

Sanja Čandrlić, Martina Ašenbrener Katić and Martina Holenko Dlab
Department of Informatics, University of Rijeka, Rijeka, Croatia
sanjac@inf.uniri.hr, masenbrener@inf.uniri.hr, mholenko@inf.uniri.hr

Abstract - This paper presents the results of a comparative research of paper-based testing versus online testing using MudRi. The research was conducted on three courses at the Department of Informatics, University of Rijeka, and included 1231 tests. In observed courses, online tests have largely replaced traditional paper-based tests in the last few years. Therefore, the aim of this research was to determine whether the online testing results differ from the results of traditional paper-based tests. The research results showed that, for the observed courses, the median values of the results achieved during MudRi online tests and traditional paper based tests are not significantly different when online tests include certain percentage of subjective test questions.

I. INTRODUCTION

An important aspect of the educational process is knowledge assessment. Knowledge assessment is required to determine the extent to which learning outcomes are achieved. The goal of knowledge assessment is not only to rate or rank students, but also to provide feedback. Therefore, in the educational process it is important to feature formative assessment as well as summative assessment. While summative assessment seeks to evaluate students' knowledge with a grade, the goal of formative assessment is to provide feedback to students about their learning in order to positively influence their learning motivation and help them to take control of their own learning [1].

Along with the development of information and communication technology (ICT), which is increasingly used in education, teachers today have computer-based testing (CBT) tools at their disposal. The popularity of such tools, especially in the sphere of higher education, stems from the fact that with the use of CBT it is possible to assess more students in less time because the computer program reviews and evaluates tests [2]. The advantages of using ICT in education have been recognized in a much broader context than just for the assessment process since the use of ICT makes possible to replace traditional classroom instructions with e-learning. Although courses could be realized entirely online by the use of web-based learning environments, a mixed or hybrid approach to e-learning is being used in practice, also called blended learning. This approach combines online learning with traditional face-to-face learning [3].

Learning Managements Systems (LMS) play a major role in the realization of e-courses. These systems enable teachers to deliver educational materials and implement collaborative activities online by the use of tools such as forums, chats and wikis. An essential part of every LMS is a subsystem that enables online testing. An online test is a test (a series of questions used to assess students' knowledge) that students solve using a computer rather than writing their responses on paper. Standard LMS systems provide various options that make the process of creating tests or quizzes very flexible. Test created in this way can be used for formative or summative assessment [4], [5]. By the use of LMS systems such as Moodle [6], Sakai [7], Dokeos [8] and Blackboard [9] it is possible to develop tests that feature different types of questions, randomly generate tests from question banks, allow students to solve the test several times, etc.

A big advantage of online testing is automated correction that makes the job easier for teachers. But, that is not the case with objective types of questions, the correction of which mainly requires teacher's involvement. In order to overcome this problem specialized modules have been developed. For example, the knowledge assessment subsystem AHyCo LMS [3], along with standard assessment provides automated correction of programming tasks. It is possible to evaluate responses automatically for questions the answer to which is code written in one of the following programming languages: C, C++, C# and SQL [10]. Tools that can be used for the evaluation of descriptive answers are being developed, e.g. Automatic Essay Assessor [11] and Intelligent Essay Assessor [12], but their use is rather complex and language-dependent.

Although assessing student's answers in case of subjective questions can be time consuming for teachers, this type of questions should not be neglected in the process of creating tests. By using subjective questions type like Essay it is possible to measure understanding of subject matter, but also students' skills in solving some types of problems and expressing ideas in written form [13].

This paper presents the analysis and comparison of results of traditional paper-based tests and online tests that students solve on a computer. The aim of the research was to determine if and what kind of online tests can replace traditional tests that students solve on paper. Based on the results, we propose a model for transition from paper-based to online testing. The paper is organized as follows. The second chapter lists the features of the system MudRi for online testing. The third chapter describes the research methodology, including the process of preparing questions

and creating tests that were used in the scope of this research. The fourth chapter presents the research results. The last, fifth chapter, presents conclusions and plans for future work.

II. TESTING WITH MUDRI

The academic community at the University of Rijeka (teachers, students and other University employees) uses the MudRi system [14]. MudRi is a customized solution for the realization of e-courses based on the open source product Moodle. The system is extensively used with the courses of the Department of Informatics. Since it offers many features in the context of online testing, MudRi has been chosen for this study. The following sections of this paper present the types of questions and features that are available while creating online tests.

A. Types of questions

MudRi offers a whole range of question types that can be present in online tests. For the purpose of this research, questions of objective type (*Multiple choices* questions, *Connection* and *True/false* questions) and questions of subjective type (*Essay* and *Short answers*) were used and are described in the remaining of this section.

The *Multiple choices* question type offers several answers of which one or more answers are accurate. It is possible to mix the order of offered answers and define the percentage of points that will be assigned for each correct answer. The *True/false* question type has only two possible answers – true and false – of which one is the correct answer and the other is the wrong answer. It is possible to define the percentage of points that each answer carries. In *Connection* questions students are supposed to connect questions or concepts on one side with answers or concepts on the other side. It is possible to mix the order of these concepts.

The *Essay* question type is used for questions that require a longer answer and that cannot be automatically evaluated through the system but the teacher must review and evaluate the response. In *Short answer* question type students enter a word or a phrase in the answer field, which is then compared with the correct answer (or more correct answers) that were entered by the teacher. It is also possible to subsequently award points for a correct answer.

When creating questions it is necessary to define a standard number of points awarded for a correct answer and a penalty factor for wrong answers. The penalty factor is an option for adaptive tests – tests that can be taken several times. The penalty factor enables the subtraction of a portion of points when the student is allowed to offer an answer to a question that was previously answered incorrectly. For each question, the teacher can enter the desired feedback that will be displayed to the student in case of correct or incorrect answers. For some question types, it is possible to define feedback for each possible answer and in this way, for example, provide detailed explanation for incorrect answers.

B. MudRi online tests

A test consists of selected questions that are stored in a question bank. Each question can be used in multiple tests. When choosing the questions that will be used in a test, the teacher can define a fixed set of questions, let the system randomly pick questions from a category or use a combined approach. In order to allow random selection of questions, it is convenient that questions are grouped by chapter, by question difficulty or by other criteria. The number of predefined points that a question carries can be modified in the process of creating a test. This means that the same questions can carry a different number of points in different tests.

When creating tests, it is possible to define a period of time during which the test is available, limit the time available for writing the test (in minutes), set the number of questions displayed on a page and set the access password. It is also possible to configure the system to shuffle the questions or answers thus reducing the possibility that a student can copy correct answers from other students during examination. The teacher can also decide whether to add a penalty for incorrect answers in order to discourage the students from guessing the correct answer. In order to prevent opening other web pages during the examination, it is possible to set up the test to be solved in a so-called 'safe window'.

Upon completion of each test, their own test results are available to students, while the teacher has access to the test results of all students. Information about each taken test is visible, such as test start and end time, time spent, number of points achieved and number of points per question. In addition, the total average of the test is calculated and the results are displayed graphically so that teachers can analyze the test as a whole and by specific questions.

III. RESEARCH METHODOLOGY

The aim of this research was to determine whether online tests can replace traditional paper-based tests and what types of questions should online tests contain. This includes an analysis of whether the online tests consisting exclusively of objective type questions can completely replace the paper-based ones, or the online test should include a certain percentage of subjective type that will be manually graded by teachers. The main hypothesis of the research is: *There is no significant difference in median values of the results achieved during MudRi online tests and traditional paper-based tests.*

The research analysed results for the tests taken on three courses at the Department of Informatics, University of Rijeka: *Data Modeling*, *Process Modeling* and *Information Systems*. These three courses offer fundamental knowledge in the field of information systems development and are suggested as important for the education of future practitioners in the field [15]. The transition to the online test was conducted in three phases performed by two teachers. The first phase included the adjustment of questions to an online test format and their insertion into the MudRi system (creating a question bank). The second phase included creation of online tests where a special attention was devoted to the composition

of the test (i.e. the ratio of subjective and objective questions type). The final, third phase refers to conducting the tests in a real environment.

A. *Creating the Question bank*

The study used questions from classic paper-based tests that were adapted to the MudRi testing environment.

Set of questions used in paper-based tests for observed courses had been created over a longer period of time. Questions used in paper-based tests were mostly *Essay* type which included description, model or picture description and listings. Questions that require a short answer or those that require to draw or describe a picture or a model were used less often. Questions of objective types were used rarely. The number of different questions used during paper-based testing for *Data Modeling*, *Process Modeling* and *Information Systems* course was 137, 62 and 149 respectively.

The transformation of questions from the previously used written format to available objective types included transformation of the question itself and the creation of possible answers. Question bank was also supplemented with certain number of subjective questions - existing *Essay* and *Short answer* questions were also translated into online form.

Table I shows the number of created questions in the online question bank, grouped by type and course. When compared to the number of questions used in paper-based tests, as stated earlier in this section, the number of different questions in online tests is significantly bigger. Besides that, the ratio between question types used in paper-based tests comparing to those created for online testing has significantly changed: in paper-based tests questions were mostly of *Essay* type while in online question bank *True/false* and *Multiple choice* questions prevail.

Questions in online tests are designed to be clear and unambiguous, not too easy and not too hard. For example, in *Multiple choice* questions it is necessary to write the question text and offer a few (one or more) correct and a few incorrect answers [13]. We tried to ensure that incorrect answers are not too easy i.e. that their identification and elimination as incorrect is not too obvious. Sometimes, it was not easy to do so. The same applies to *True/false* question type.

Compared to conventional paper-based tests, online tests are restricted when it comes to the use of questions that need to be answered by a drawing (e.g. drawing a diagram, graph or model). This type of questions is not supported by online tests. Therefore, apart from online tests, other assessment activities were used to estimate students' knowledge.

TABLE I. ONLINE TESTS QUESTION BANK

Category of questions	Data Modeling	Process Modeling	Information Systems
Essay	25	26	0
Short answer	3	0	0
Connection	31	9	50
Other: True /false, Multiple choice	159	184	540
Sum	218	219	590

Since inserting questions and answers into the MudRi system is a time-consuming task, we considered automatic import from a file offered by the system. MudRi offers the ability to import questions that were created in other solutions such as Blackboard and Hot Potatoes and import of files of the following formats [14]: Aiken, GIFT, Moodle XML, etc.

We prepared questions for each chapter in a textual document, and therefore chosen the Aiken format [16] to import them into the MudRi system. The format allows the creation of *True/false* and *Multiple choice* questions and these two types of questions were the most common in our question bank. The Aiken format is quite simple. The question is formulated in the first paragraph, and every subsequent paragraph provides a response. The proposed answers begin with a letter which is followed by a dot '.' or a bracket ')', and space ' '. The list of answers is followed by a paragraph that begins with the keyword "ANSWER:" and this is followed by a letter that indicates the correct answer. In case that a single file contains several questions, they have to be separated by an empty paragraph.

Essay, *Connection* and *Short answer* types were manually inserted.

During test grading, the points assigned to student's correct answers are summed up. For the process of grading *Multiple choice* questions that have multiple correct answers, we used the rule that an incorrect answer nullifies a correct answer in order to discourage students from guessing the correct answer. For example, if we have a task with five answers from which two are correct, each of the correct answers carries 50% of the question points (the sum is 100%), and each of the incorrect answers - 33.33% (the sum is 100%). This system ensures that in case that a student picks all answers, the sum will be 0% (i.e. no points will be awarded).

B. *Creating the tests*

Online tests were set up to be solved within a 'safe window' and in a specific timeframe. Each test had an access password defined which was given to students just before the test begins. Tests consist of 15 – 25 questions from different categories. For every online test a part of questions was predetermined and a part was randomly chosen by the system from specified categories in order to cover all course lessons.

Traditional paper-based tests included *Essay* type of questions, questions for which it was required to draw a sketch, and *Short answer* questions.

For the purpose of this study, tests for different courses were designed differently, in respect to the question types used. In the *Data Modeling* course and the *Process Modeling* course online tests one third of questions was subjective (*Essay* and *Short answer* types) and two thirds were objective (*Multiple choice*, *True/false*

TABLE II. TEST PROPERTIES

	Classic paper-based tests	Online tests	Online tests
	All three courses	Data Modeling, Process Modeling	Information Systems
Question type	Essay type of questions, short essays and listings, image drawing, short answer questions	Multiple choice, true/false questions, connection; essay and short answer	Multiple choice, true/false questions, connection;
Duration	60 minutes (paper-based)	20 minutes (computer-based)	20 minutes (computer-based)
Number of questions	15 (on average)	20 (on average)	25 (on average)
Number of tests	753	310	168

questions, *Connection* question types). Objective questions carried fewer points than *Essay* or *Short answer* type of questions. For the *Data Modeling* and the *Process Modeling* course points based on answers for objective type of questions was 50%, i.e. almost 70% of objective type questions carried 50% of points and 30% of subjective type questions carried 50% of points. Questions in online tests for the *Information Systems* course were only objective. Test properties are compared in Table II.

After students hand in the test for evaluation, they are presented with their test results, correct answers, feedback related to particular questions and overall feedback related to the whole test and scoring. *Essay* and *Short answer* questions are graded later manually by the teacher. After closing the test, points and overall feedback is visible to the student.

C. Participants and online testing properties

The research was held at the Department of Informatics, University of Rijeka, so the participants were students of Single major and Double major program of informatics at the Department (second and third year of undergraduate study).

We have started with online testing in 2009 in the *Data modeling* and the *Process modeling* course, using tests created as described in previous section. After two years of using online tests with combination of subjective and objective question types, our intention was to ease the correction of tests and therefore use only objective question types in the *Information systems* course. The motivation for such approach arises from the fact that in recent years the number of students, and consequently teachers' teaching load, has increased. This is especially the case when teachers hold several courses with two or three test planed.

Both, online and paper-based tests were solved in the classroom under controlled conditions. In order to solve the traditional test students had 60 minutes, and in this time, they had to solve about 15 questions. Although online tests had more questions than paper-based tests, less time was needed to solve them due to different questions types.

Grading traditional paper-based tests required significantly more time than grading online tests. Online tests were graded automatically by the system immediately after students submitted them. Teachers manually graded *Essay* type questions (if present in the test) and given points are automatically added to the total sum of points.

All the tests carried through the study were used as part of the summative assessment. Minimum of 50% test points was necessary for passing the test. Apart from online and paper-based tests, teachers used other activities in order to assess students' knowledge in mentioned courses.

IV. COMPARISON OF TEST RESULTS AND DISCUSSION

The goal of the research was to determine whether there was a significant difference between median values of the results achieved during MudRi online tests and traditional paper-based tests. The study included 753 paper-based and 478 online tests. In the *Data modeling* and in the *Process modeling* course used tests included both, subjective and objective types of question, while those for the *Information systems* course consisted only of objective ones. Descriptive statistical analysis for all three courses was performed. The results are presented in Table III.

High variance coefficient for paper-based test results shows that arithmetic mean cannot be considered as representative central tendency indicator. Rather low value of quartile deviation coefficient refers to median as better indicator of central tendency.

It is also important to notice that measures of variance are significantly lower for online tests. Based on the Levene's test for equality of variances, in all three cases the hypothesis of equality of variances can be rejected with extremely high significance ($p<0.0001$), i.e. we can conclude that variances for online tests and paper-based ones significantly differ.

The tests results for *Data Modeling, Process Modeling* and *Information Systems* courses are shown using the Box-Plot graph in Fig. 1, 2 and 3 respectively. In each figure the results for online test are shown in the left part of the graph, and for paper-based test in the right part of the graph.

TABLE III. DESCRIPTIVE STATISTICAL ANALYSIS

	Data Modeling		Process Modeling		Information Systems	
	Online tests	Paper-based tests	Online tests	Paper-based tests	Online tests	Paper-based tests
N	162	226	148	254	168	273
Arithmetic mean	0.6475401	0.6025221	0.5643795	0.5626378	0.6349981	0.5535165
Range	0.2415-0.9635	0-1	0.221-0.8915	0-1	0.33-0.93042	0-1
Standard deviation	0.140849	0.240482	0.135689	0.249787	0.136961	0.262828
Variation coefficient	0.2175147	0.3991249	0.2404207	0.4439565	0.2156873	0.4748339
Median	0.65875	0.63	0.56275	0.58	0.634095	0.57
Interquartile range	0.2125	0.27	0.1625	0.32	0.202145	0.36
Quartile deviation coefficient	0.164674	0.212598	0.146961	0.283843	0.160266	0.321586

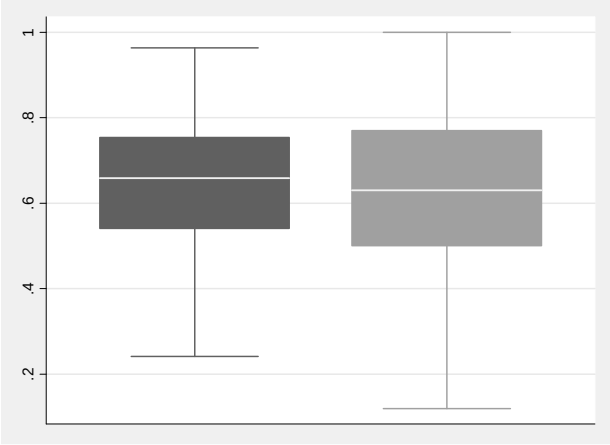

Figure 1. Box-plot diagrams for Data Modeling course (left online tests, right paper-based tests)

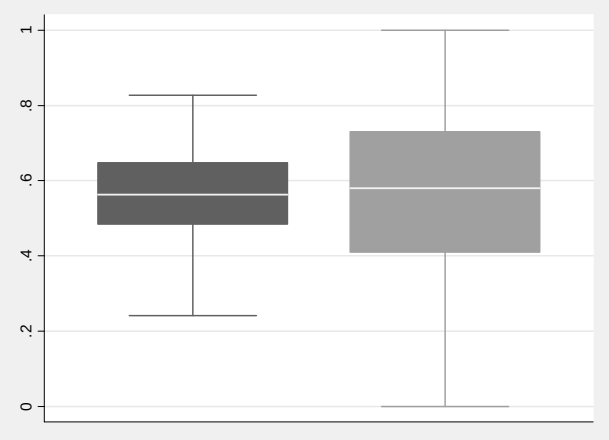

Figure 2. Box-plot diagrams for Process Modeling course (left online tests, right paper-based tests)

Figure 3. Box-plot diagrams for Information Systems course (left online tests, right paper-based tests)

Online tests, in which the majority of questions are of the *Multiple choice* or *True/false* type exclude the impact of subjective factors related to the teacher as the evaluator, which were potentially present while correcting *Essay* type of questions in paper-based tests.

Normality analysis (Table IV) indicated the use of nonparametric statistic methods for hypothesis testing. By using Mann-Whitney test it is stated that for the *Data Modeling* course and the *Process Modeling* course there is no significant statistic difference in median values (p=0.2401 and p= 0.3777, respectively) between online and paper-based tests. For the *Information Systems* course the difference is extremely significant (p=0.0066).

TABLE IV. NORMALITY ANAYSIS

	Data Modeling		Process Modeling		Information Systems	
	Online tests	Paper-based tests	Online tests	Paper-based tests	Online tests	Paper-based tests
Shapiro-Francia W'	0.98905	0.94746	0.99238	0.97006	0.99270	0.96782
P	0.20769	0.00001	0.53225	0.00009	0.47966	0.00002

Therefore, for the *Data Modeling* and the *Process Modeling* the hypothesis is confirmed – median values do not differ significantly. For the course *Information Systems* whose tests consist only of *Multiple choice*, *True/false* and *Connection question* types (i.e. objective questions), the difference in median values between online and paper-based test results is significant. Better results on online tests in the case of the *Information Systems* course can be explained by the fact that *True/false* and *Multiple-choice* questions can be answered even with lower levels of knowledge and based on recognition.

A large difference in the maximum number of points achieved in paper-based tests and in online tests can be noted. We conclude that the subjective factor of the examiner plays an important role in the assessment of descriptive answers, since the examiner has the possibility to assess the answer as partially correct, or to assess the "hidden knowledge" in the given answer. Differences can also be observed for the minimum number of achieved points in paper-based tests and online tests. On written tests, students will very often give up from finishing the test if they estimate that they will not be able to achieve the enough number of points. This is the reason why we recorded the minimum score of 0 in the analysis of paper-based tests.

Based on the result of the study, we conclude that online tests can replace traditional paper-based tests for student' knowledge assessment, but special attention should be paid to its composition. Therefore, we propose a following model for transition to the online testing.

In the phase of questions creating, existing question bank needs to be adapted to the online testing environment. The certain number of existing *Essay* type questions could be transformed to the questions of objective type (e.g. *Multiple-Choice*, *Connection*), while some of them could be just imported in online system as they are. Depending on the course domain, question types offered by the chosen system can or can not be adequate. In case there is no adequate question type (like drawing questions in our case), additional assessment activities can be performed [17]. When creating *Multiple-Choice* questions with multiple answers, we suggest to use the rule that an incorrect answer in certain percentage nullifies a correct answer. In addition, we highly recommend entering questions into the system using the Aiken or similar format because it is faster than entering questions

manually. Only minor modifications and manual input of questions that the format does not support are necessary.

Concerning the test creating phase, we consider that it is best to use a combination of online tests and other methods of knowledge assessment. In order to prove that they have adopted the course material, students have be able to explain what they learned, which is possible with the *Essay* type of questions. Therefore, instead of online tests in which the questions are posed in the *True/false* or *Multiple choice* types, we propose that online tests have around 30% of subjective questions (*Essay, Short Answer*). Although it is not possible to correct them automatically, this ratio does not represent great time workload for the teacher (it's about 2-3 questions per test).

V. CONCLUSION AND FUTURE WORK

This paper presents the results of a research that included 1231 paper-based and online tests. If the results are compared with the results of online tests, we can conclude that it is possible to replace paper-based tests with online tests in order to assess student knowledge. The biggest advantage of paper-based on the other hand is the short time in which it is possible to prepare the test and the possibility to ask questions that require a drawn response.

Based on students' comments that we received during the testing activities, we conclude that they prefer to use the keyboard rather than a pencil so they would rather take online tests than paper-based tests.

Due to above-mentioned, we plan to use them in the future in the scope of observed courses. Our future work will include efforts to further improve methods for online assessment. This includes revising the question bank. Students responses will be analyzed in order to identify to easy questions (the questions with 100% correct answers), as well as questions for which there are no correct answers (the questions that are potentially unclearly stated and therefore incomprehensible to the students). We also plan to update the question bank with sets of different types of questions for formative assessment. Students will be able to solve a self-assessment test after completing each course unit. This will enable them to test their knowledge, in the process of preparation for the tests that follow.

VI. ACKNOWLEDGEMENTS

The research has been conducted under the project "Extending the information system development methodology with artificial intelligence methods" (reference number 13.13.1.2.01) supported by University of Rijeka (Croatia).

REFERENCES

[1] D.J. Nicol, and D. Macfarlane-Dick, Formative assessment and self-regulated learning: a model and seven principles of good feedback practice, Studies in Higher Education, 2006; 31(2): pp. 199-218.

[2] S. M. Sidhu, Computer-Based Assessment. In Tomei, L. A. (Ed.), Encyclopedia of Information Technology Curriculum Integration, 2008; pp. 131-137.

[3] N. Hoic-Bozic, V. Mornar, and I. Boticki, Introducing adaptivity and collaborative support into a Web-based LMS, Computing and informatics. 2008; 27(4): pp. 639-669.

[4] J. P. Sewell, K. H. Frith, and M. M. Colvin, Online Assessment Strategies: A Primer, 2010; 6(1): pp. 297-305.

[5] J. Cole and H. Foster, UsingMoodle - Teaching with the Popular Open Source Course Management System, 2nd edition, O'Reilly Media Inc., 2008.

[6] Moodle.org: open-source community-based tools for learning, http://moodle.org (accessed on October 19, 2011)

[7] Sakai Project: collaboration and learning - for educators by educators, http://sakaiproject.org (accessed on October 19, 2011)

[8] Dokeos: Open Source E-learning, http://www.dokeos.com (accessed on October 19, 2011)

[9] Blackboard: Technology and Solutions Built for Education, www.blackboard.com (accessed on October 19, 2011)

[10] N. Hoic-Bozic, I. Budiscak and I. Boticki, Online assessment of programming assignments in a learning managment system AHyCo, Engineering Review. 2008; 28(1), pp. 51-60.

[11] T. Kakkonen and E. Sutinen, Evaluation Criteria for Automatic Essay Assessment Systems – There is much more to it than just the correlation, Proceedings of the 16th International Conference on Computers in Education, 2008.

[12] L. Streeter, J. Pstoka, D. Laham and D. MacCuish, The credible grading machine: Automated essay scoring in the dod. In Proceedings of Interservice/Industry, Simulation and Education Conference (I/ITSEC), 2003.

[13] A. W. Ward and M. Murray-Ward, Assessment in the Classroom. Belmont, CA: Wadsworth Publishing Co., 1999.

[14] MudRi: sustav za učenje Sveučilišta u Rijeci, http://mudri.uniri.hr (accessed on October 19, 2011)

[15] M. Pavlic, M. Marinovic and S. Candrlić, Model of Basic Training for Information Systems Developers, Proceedings of the 8th International Symposium on Operational Research in Slovenia SOR'05; Ljubljana: Slovenian Society Informatika - Section for Operational Research (2005); pp. 411-417.

[16] P. McElroy, D. Brabazon and E. McLoughlin, Analysis of use of multiple choice question (MCQ) examinations as an assessment tool for postgraduate engineering modules. Proceedings of the Irish Learning Technology Association Conference, IT Sligo. 2006.

[17] N.Hoic-Bozic, M. Holenko Dlab, and E. Kusen, A Blended Learning Model for "Multimedia Systems" Course. In Workshop on Learning Technology for Education in Cloud (LTEC'12), Springer Berlin Heidelberg, 2012, pp. 65-75.