Comparing the Effectiveness of Learning Content Management Systems to Intelligent Tutoring Systems

Ani Grubišić, Slavomir Stankov, Zdeslav Hrepic

Abstract — The educational utility of instructional software should be thoroughly evaluated during early implementations in the educational process. In this paper, we present the results of an experiment in which we compared a student's content learning with an online course management system, the Blackboard™, and with a representative of web-based authoring shells for building intelligent tutoring systems, xTEx-Sys. The experiment was coordinated remotely from a distant location (another continent) by the xTEx-Sys developers and was carried out by a proficient Blackboard™ user. We compared student learning with these two instructional e-learning systems by using the effect size as the primary metric. This experiment involved English speaking students accustomed to working with Blackboard™ (xTEx-Sys was previously used by Croatian-speaking students only). The instructor who taught all student groups involved in the experiment had no part in designing either xTEx-Sys or Blackboard™. The results gained through this experiment have shown us that there was no statistically significant difference in using these two systems, but the surveys have revealed some difficulties in using xTEx-Sys.

Index Terms — Effectiveness, E-learning, Evaluation, Intelligent tutoring systems

1 Introduction

Information and communication technologies (ICT) have become an integral part of educational systems as a support for teachers in the learning and teaching process. An intersection between the world of information and communication technology and the world of education is nowadays known as e-learning and it is enabled by the e-learning systems [1].

It is generally accepted that instructional software should be evaluated before being used in the educational process on a wider scale. In addition to gauging the software's utility, such evaluation is critical for obtaining early insights into different and innovative ways in which the respective technology package can or cannot be used to support the learning and teaching process. A well-designed evaluation should provide the evidence whether a specific approach has been successful and whether it could be of potential value to other instructors [2].

There are different kinds of e-learning systems. In this paper we are particularly interested in two rather different types of these systems: intelligent tutoring systems (ITS) and learning content management systems (LCMS).

Intelligent tutoring systems are computer systems that support and improve the learning and teaching processes in a variety of knowledge domains, while respecting the individuality of the learner, as is the case in traditional "one-to-one" tutoring ([3], [4], [5]). Tutoring has been confirmed to be successful and presents the most efficient learning and teaching process ([6], [7]). The major problem associated with ITS is their expensive and time consuming development process. In order to overcome these problems we chose an approach of creating a particular ITS from flexible shells acting as program generators.

Learning content management systems (LCMS) are computer systems used to create, edit, manage, and publish educational contents. There are dozens of LCMS available. One of the most popular LCMS is Blackboard™ (www.blackboard.com), which is used by thousands of educators around the world [8].

Blackboard™ allows instructors to disseminate materials and readings to students, to create online tests, maintain gradebooks, organize discussion boards, etc. Blackboard™ is an online course system deployed on international scale. It can be used to administer online multiple choice quizzes, as well as a depository site for

- *A. Grubišić is with the Department of Computer Science, Faculty of Science, Split, Croatia. E-mail: ani.grubisic@pmfst.hr.*
- *S. Stankov is with the Department of Computer Science, Faculty of Science, Split, Croatia. E-mail: slavomir.stankov@pmfst.hr.*
- *Z. Hrepic is with the Department of Physics, Fort Hays State University. E-mail: zhrepic@fhsu.edu.*

course-related materials and presentations. Simple and powerful web-based tools allow teachers to build and manage learning content and provide an attractive environment for students. Synchronous and asynchronous tools support student interaction, small-group work, and peer knowledge sharing. Tests, quizzes and assignments are easy to create and deploy, and provide timely feedback and reporting for students [9].

In this paper, we present the results of an experiment in which we compared the learning of students who used the Blackboard™ system, with the learning of students who used the eXtended Tutor-Expert System (xTEx-Sys) [10]. The latter system is representative of web-based authoring shells for building intelligent tutoring systems (ITS). In this experiment the xTEx-Sys system has been used by English speaking students for the first time (in previous experiments the xTEx-Sys system was used only by Croatian language speakers).

Before presenting the results of the current experiment, we reviewed the fifteen-year long research and development of the Tutor-Expert System (TEx-Sys) model [11], and we gave an overview of the evaluation methodology that has been used in this research.

2 BACKGROUND

The first implementation of an intelligent authoring shell model called the TEx-Sys was the on-site TEx-Sys (1992-2001) [11]. It has enabled the realization of the learning and teaching process for a student and a teacher, enabling them their basic functionalities respectively: (i) designing learning contents related to any domain of knowledge and (ii) learning and teaching as well as testing and evaluating knowledge. After that followed the web-based intelligent authoring shell (1999-2003, Distributed Tutor-Expert System, DTEx-Sys) [12], the system based on the dynamic web documents. This web-oriented ITS has been developed by keeping in mind issues like (i) accessibility for a large number of potential users and (ii) learning and teaching in arbitrary domains. Finally, the system based on web services (2003-2005, xTEx-Sys) [10] was implemented where we strictly differentiate the functionalities between experts and teachers.

The xTEx-Sys system is a web-based authoring shell with an environment that can be utilized by the following users: an expert who designs the domain knowledge base, a teacher who designs courseware for the student learning and teaching process (as well as tests for the student knowledge evaluation), a student who selects a course and navigates through the domain knowledge content using the didactically prepared course content, and finally, an administrator who supervises the system.

The xTEx-Sys system is not only a quiz-administering tool, but it also provides a practice venue for students. In addition, questions that xTEx-Sys administers are not predefined, but rather generated by the system based on the predefined semantic network structures related to the domain of interest.

In the past decade, there were numerous applications of the TEx-Sys model in the learning and teaching process that involved students in all educational levels, from primary to tertiary. In the period from 2001 to 2007, 1302 students took 5482 content knowledge tests in one of the TEx-Sys model versions in order to evaluate their understanding of different knowledge domains.

Questionnaires about the students' impressions were given to the students after finishing the courses that were supported by the TEx-Sys model. The qualitative analysis of the questionnaires' results revealed that most students expressed positive attitudes toward using the model, and they were open to idea of embracing that kind of learning and teaching support. The qualitative analysis alone was not sufficient to determine the utility of the TEx-Sys system. Therefore, we conducted 11 quantitative experiments in order to evaluate the educational influence of the TEx-Sys model [13].

3 EVALUATION METHODOLOGY

The previously mentioned 11 quantitative experiments used a modification of a classical two-group experimental design with pre-and-post tests [14]. The majority of them (8 experiments) involved primary school pupils. It is very interesting to note that primary school pupils have embraced the learning and teaching process with the xTEx-Sys system more readily than university students, what can be seen while observing calculated effect sizes and survey results. Results gained through the conducted experiments have shown a need for adding some extended functions for courseware development and learning management in the xTEx-Sys system in order to get it as close as possible to the Bloom's 2-sigma target [7].

In the current experiment, which compared the xTEx-Sys system to Blackboard™ effectiveness, we deployed a factorial

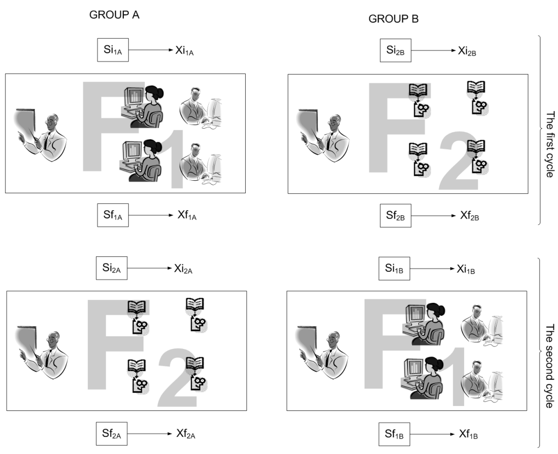

Fig. 1. A factorial design

experimental design [15]. In a factorial design we have two or more statistically equivalent parallel groups in which factors are introduced by rotation in each cycle.

At the beginning of each cycle (Fig. 1.), initial states Si_{1A}, Si_{2A}, Si_{2B} and Si_{1B} and their respective evaluation means Xi_{1A}, Xi_{2A}, Xi_{2B} and Xi_{1B}, were established using a pre-test before introducing the experimental factors. Pre-tests are used because of differences in initial students' skills and backgrounds.

At a later time, students took the comparable test in order to determine the knowledge improvement accomplished by the educational intervention. In the first cycle, group A used the xTEx-Sys system (experimental factor F_1) and the group B used Blackboard^{TM} (experimental factor F_2). In the second cycle the groups rotated systems so group A used Blackboard^{TM} (factor F_2) and group B used the xTEx-Sys system (factor F_1). At the end of each cycle, final states Sf_{1A}, Sf_{2A}, Sf_{2B} and Sf_{1B} and respectively their means Xf_{1A}, Xf_{2A}, Xf_{2B} and Xf_{1B}, were determined using a post-test, in order to calculate the difference between Blackboard^{TM} and the xTEx-Sys system effectiveness. Therefore, in the first cycle the experimental group was group A, and in the second, the experimental group was group B.

In a factorial design, different domain knowledge is learned in each cycle. Thus, superiority of the method associated with one factor can be established if that method is found to be better in each cycle, regardless of the groups' equivalence and the content taught.

4 DESCRIPTION OF THE EXPERIMENT

The experiment took place at Fort Hays State University (FHSU), a state supported University in Kansas, USA. The experimental design involved four (4) sections of Physical Science Lab course (PHYS 103) for non-science majors [16], all taught by the same instructor. The course meets once per week for two hours and course assessment traditionally consists of three components:

1. Pre-lab quizzes (pre-labs), which students take individually online and before the lab. Pre-labs are based on a short introductory reading related to the lab content and serve to familiarize students with the topic.
2. Lab reports, the second assessment component which students complete collaboratively during class time.
3. Tests associated with lab material, which typically involves short answer and multiple-choice questions. Tests are administered as paper-and-pencil, in-class exams.

One of the authors taught this course each semester between Fall 2004 and Spring 2008, and in this entire period the tests were the weakest assessment component of the students' performance in the course. Another three instructors taught one or more sections of the course in the same period, and this entire time the test results hardly depended on the instructor teaching the course.

In order to remedy the situation, in Fall 2007 we deployed post-lab quizzes (post-labs) which were also taken online after completion of the lab. Post-labs serve as lab material review and reinforcement. Post-lab quizzes, as well as pre-lab quizzes consisted of pre-defined multiple choice questions and were administered using Blackboard^{TM}. The success with implementation of post-lab quizzes in Fall 2007 was limited.

A semester later, in Spring 2008--in an attempt to possibly improve the effect of post-lab quizzes--we administered them using the xTEx-Sys system, which does not use predefined questions, but rather questions dynamically generated by the system. At the same time, we used the Blackboard^{TM} administered post-lab quizzes with pre-defined questions as a control for the xTEx-Sys system effect on students' learning.

Our research question was whether or not administration of post-lab tests on the xTEx-Sys system affects, i.e. improves, student learning differently than the Blackboard^{TM} administered quizzes, as measured by students' test scores.

We have randomly assigned two (out of four) sections to the experimental group and the other two to the control group. The experimental treatment consisted of taking post-lab quizzes related to three labs on the xTEx-Sys system while the control group was taking content-wise corresponding post-lab quizzes on the Blackboard™ system. The experimental group also had the opportunity to learn the content using the xTEx-Sys learning and teaching features. All other parts of the course operated identically for students in all sections.

After completion of the first three labs involved in the experiment, all sections took the same in-class test (the checkpoint test CHK1), after which the control and experimental sections exchanged. After another three labs, all sections again took the same in-class test--this time the checkpoint test CHK2. Each test covered only the content related to the three labs that preceded it.

Upon completion of the second test, we gauged students' attitudes toward the xTEx-Sys and the Blackboard™ systems using an online administered questionnaire.

4.1 Subjects

Fort Hays State University, where the experiment was carried out, is a public, open admission university in the state of Kansas, USA. Students who took the physical science lab course involved in the experiment were non-science majors who typically take the course for their general education requirements during their freshman or sophomore year.

The experiment started in February 2007 and lasted until April 2007 (for total of 10 weeks). Out of 65 students initially enrolled in the course, 48 (74%) completed both the pre- and post-test part of both tests (CHK1 and CHK2) involved in the experiment. Others who did not complete all these parts either dropped the course or simply did not take one or more of the four tests (either pre-or post).

Out of the initial 65 participants, 26 students were assigned to the experimental group A (lab sections A and D) and 39 students to the experimental group B (lab sections B and C). Group A was the experimental group in the first cycle and Group B was the experimental group in the second cycle. The assignment of each of the lab sections into experimental and control groups was random.

4.2 Procedure

The experiment was conducted according to the schedule presented in Table 1. During the first lab session students were introduced to the course and were given a short overview of procedures associated with the experiment. The first session was concluded with the first

TABLE 1

THE EXPERIMENT SESSION SCHEDULE

	Group A	Group B
Session 1	45 min	45 min
Pre-test	45 min	45 min
Session 2	5 weeks	5 weeks
Learning and teaching process – part 1	2 hours/week	2 hours/week
Checkpoint test 1	45 min	45 min
Session 3	5 weeks	5 weeks
Learning and teaching process – part 2	2 hours/week	2 hours/week
Checkpoint test 2	45 min	45 min
Survey	15 min	15 min
Total	**10 weeks**	**10 weeks**

pre-test. As mentioned earlier, the experiment involved two learning-test cycles, each followed by an associated checkpoint test (CHK1 and CHK2).

After completion of the second checkpoint test, students' attitudes toward each of the systems were gauged using an online survey. The survey questions (shown in Fig 2) targeted students' perceptions of the training quality. Students answered each question by selecting their level of (dis)agreement on a five-point scale ranging from strongly agree (5) to strongly disagree (1) with a neutral option (3) included.

Time allocated for online survey completion, as well as for completion of all other procedures involved in this experiment, was identical for the experimental and control groups.

4.3 Data Analysis

While analyzing results it was important to determine the magnitude of the student drop-off from each group. By the end of the experiment, all pre- and post-tests were completed by 21 out of 26 students in Group A, (81%) and by 27 out of 39 students in Group B (69%).

T1 It was easy to use the xTEx-Sys system:

T2 Browsing concept definitions in xTEx-Sys system helped me better understand the content:

T3 Taking post-lab quizzes in xTEx-Sys system helped me better understand the content:

T4 It was easy to understand the post-lab quiz questions in xTEx-Sys system:

T5 It was easy to understand the answer options of the post-lab quiz questions in xTEx-Sys system:

T6 Difficulty level of the post-lab quiz questions in xTEx-Sys system was very high:

T7 I would recommend using xTEx-Sys system POST-LAB QUIZZES in physical science lab course in future:

T8 I would recommend using xTEx-Sys system for studying and reviewing in other courses in future:

Fig. 2. The survey questions related to the xTEx-Sys (the same ones were used for the Blackboard™)

A standard significance test was used to determine the effect of different procedures on the dependent variable. First, we checked whether groups' initial competencies were equivalent before comparing the gains of the groups. That meant calculating the means of the pre-test scores for both groups and their standard error of mean.

The null-hypothesis H0 was stated for each checkpoint-test: "There is no significant difference between Group A and Group B in the results of test $H0_{CHK1}$ and ($H0_{CHK2}$) respectively".

After determining the gain scores associated with each test, we used the t-test (after testing if the variables have normal distribution) to determine the statistical significance of the difference between the groups for every checkpoint test ($\alpha = 0.05$, two-tailed).

4.4 Results

Table 2 contains the descriptive statistics. The columns "Pre-test", "CHK1" and "CHK2" show the calculated values for mean, median, and standard deviation of the raw data collected during the pre-test, first checkpoint test and second checkpoint test for both groups.

The columns of Table 2. that start with "Gain" show the calculated values for mean, median, and standard deviation of the differences between post-test, first checkpoint test and second checkpoint test. The differences between CHK1, CHK2 and respective pre-test score means are positive and almost identical for both groups.

A prerequisite for applying the t-test is the assumption of normal distribution of the variables in the test samples. A Kolmogorov-Smirnov test was used for normality testing (Table 3.). The probability of the K-S coefficient for every test is larger than α=0,05. Therefore, the sample as a whole and the groups themselves, have normal distribution, and the t-test--for determining the existence of statistically significant difference among the groups--can be performed.

Table 4. shows the results of testing hypothesis H0 using a two-tailed t-test for an independent sample. Effect sizes are calculated as Glass's Δ [17] using the following equations:

$$\Delta_{CHK1} = \frac{\overline{X}_{A_{CHK1}} - \overline{X}_{B_{CHK1}}}{\sigma_{B_{CHK1}}} = \frac{5{,}81 - 5{,}59}{4{,}53} = 0{,}05$$

$$\Delta_{CHK2} = \frac{\overline{X}_{B_{CHK2}} - \overline{X}_{A_{CHK2}}}{\sigma_{A_{tCHK}}} = \frac{3{,}67 - 3{,}67}{4{,}85} = 0{,}00 \quad (1)$$

Calculated effect sizes are, according to Cohen [18], very small. The first column of Table 4. specifies the test, the second column represents the effect size, the third column the degrees of freedom, the fourth column the t-value of the study, and the fifth column provides the associated p-value.

By examining columns four and five of Table 4., one can see that the experimental groups in both cycles have not achieved a statistically and practically significant result for the dependent variable, that is, student's knowledge.

Context information about the participants was collected in a survey. Out of 65 students initially enrolled in all four sections, a limited number (28 i.e. 43%) took the survey. This sub-population consisted of four males and 24 females, mostly in their freshman (16) and sophomore (5) years. Virtually all of them rated their computer usage skills as either excellent (8) or good (18) as opposed to poor

TABLE 2

DESCRIPTIVE STATISTICS

	Pre-test	CHK1	CHK2	Gain CHK1 and Pre-test	Gain CHK2 and Pre-test
Group A (21 students)					
Mean	20,19	26,00	23,86	5,81	3,67
Stdev.	4,9155	5,13	6,7103	4,18	4,85
Group B (27 students)					
Mean	17,81	23,41	21,48	5,59	3,67
Stdev.	3,3171	4,57	6,3753	4,53	5,97

TABLE 3

THE KOLMOGOROV-SMIRNOV TEST OF NORMALITY

One-Sample Kolmogorov-Smirnov Test				
Both groups		Pre-test	CHK1	CHK2
N		48	48	48
Mean		52,29	50,79	40,06
Std. Deviation		10,468	15,250	8,936
Most Extreme Differences	Absolute	,123	0,109	0,149
	Positive	,123	0,109	0,149
	Negative	-,103	-0,059	-0,088
Kolmogorov-Smirnov Z		1,030	0,855	0,758
Asymp. Sig. (2-tailed)		0,240	0,458	0,613
Group A		Pre-test	CHK1	CHK2
N		21	21	21
Mean		55,43	55,43	53,90
Std. Deviation		10,787	10,787	15,623
Most Extreme Differences	Absolute	0,140	0,140	0,113
	Positive	0,135	0,135	0,113
	Negative	-0,140	-0,140	-0,091
Kolmogorov-Smirnov Z		0,643	0,643	0,519
Asymp. Sig. (2-tailed)		0,803	0,803	0,950
Group B		Pre-test	CHK1	CHK2
N		27	27	27
Mean		48,37	49,85	48,37
Std. Deviation		14,789	9,718	14,789
Most Extreme Differences	Absolute	0,172	0,128	0,172
	Positive	0,172	0,128	0,172
	Negative	-0,077	-0,125	-0,077
Kolmogorov-Smirnov Z		0,892	0,663	0,892
Asymp. Sig. (2-tailed)		0,403	0,772	0,403

TABLE 4

THE HYPOTHESIS TESTING

	Effect size Δ	df	t-Value	p-Value
CHK1	0,05	46	1,880	0,066
CHK2	*0,00*	*46*	*1,255*	*0,216*

(2). The survey questions were outlined earlier in Fig. 2. “Agree and Strongly Agree”, as well as, “Disagree and Strongly Disagree” answer categories are grouped in Table 5. The remainder up to a hundred percent for each of the systems is the percentage of neutral choices.

The comparison of students’ answers related to each of the systems shows considerable dissatisfaction with the xTEx-Sys system and preference toward the Blackboard™ system. In addition to problems with understanding the semantics used in the xTEx-Sys system, 18 out of 24 students reported problems with logging into the xTEx-Sys system, while none of them had such problems with Blackboard™. Four out of 24 survey respondents never actually managed to log into the xTEx-Sys system and were automatically excluded from the survey portion related to the xTEx-Sys system.

All survey participants were given an option to make comments or improvement suggestions, as well as to raise issues and describe problems that they encountered during any of the treatments. With respect to the xTEx-Sys system, 12 students used this option and gave their written input. Although some of these comments were positive, most of them were not.

Negative comments and improvement suggestions were mainly associated with the technical aspects of the system usage and the difficulty of understanding the structure of the domain knowledge that is based on semantic network with frames. Complaints were mostly related to difficulty in understanding the presented information and understanding the wording of questions asked. At the same time 7 students wrote comments related to the Blackboard™ system. These comments were mostly positive and frequently expressed preference for Blackboard™ when compared to the xTEx-Sys system.

TABLE 5

THE SURVEY RESULTS

	xTEx-Sys		Blackboard™	
	Agree and Strongly Agree	Disagree and Strongly Disagree	Agree and Strongly Agree	Disagree and Strongly Disagree
It was easy to use the system	12.5%	62.5%	96.4%	3.6%
Browsing concept definitions in xTEx-Sys system helped me better understand the content	12.5%	66.7%	NA	NA
Taking post-lab quizzes helped me better understand the content	20.8%	70.8%	53.6%	14.3%
It was easy to understand the post-lab quiz questions	12.5%	83.3%	74.1%	3.7%
It was easy to understand the answer options of the post-lab quiz	16.7%	70.8%	82.1%	7.1%
Difficulty level of the post-lab quiz questions was very high	41.7%	29.2%	10.7%	21.4%
I would recommend using the system's post-lab quizzes in this course in future	8.3%	75.0%	67.9%	7.1%
I would recommend using the system for studying and reviewing in other courses in future	12.5%	66.7%	60.7%	10.7%

4.5 Interpretation of Results and Discussion

At the end, we summarized the results of the experiment with regard to null hypothesis H0 in Table 6. Statistical significance (stat. sig.), mentioned in that table means that null hypothesis could be rejected at significance level α=0.05. Practical significance (pract. sig.) means that null hypothesis could not be rejected, but effect size is $\Delta \geq 0.5$. If statistical significance is achieved, practical significance is not mentioned. Positive effect (+) means that no significance is not mentioned. Positive effect (+) means that no practical significance could be observed but effect size is $\Delta > 0$. No effect or negative effect (-) means that effect size is $\Delta \leq 0$.

Regarding the first checkpoint test and $H0_{CHK1}$, the expected positive learning effect of the xTEx-Sys system could be observed, but it was statistically insignificant. In other words, the experimental group (Group A – xTEx-Sys) performed better than the control group (Group B - Blackboard™), but it was not statistically significant.

Regarding the second checkpoint test and $H0_{CHK2}$, the expected positive learning effect was not observed. In other words, neither the control (Group A - Blackboard™) nor the experimental group (Group B - xTEx-Sys) was statistically significantly better. If we compare the test results means presented in Table 2., we can observe that Group A had a larger mean than Group B in all conducted tests, regardless of the system it used.

The weak or no effect observed when comparing the performance of experimental to control groups in each cycle of the experiment

TABLE 6

THE EXPERIMENT RESULTS SUMMARY

Experimental group vs. Control group	**Dependent variable – student knowledge**	
	Statistical significance / Practical significance	Positive effect size / Negative effect size
CHK1	none	+
CHK2	none	undefined

can probably be attributed to the inclusion of the xTEx-Sys system in the treatments of the experimental groups. Therefore, the xTEx-Sys system has not shown any greater effect on the students' learning outcomes in this experiment when compared to Blackboard™. That is, we can use either of the two systems because we will get the same results.

In any case, it should be emphasized that the presented exploratory research is just the first step of a series of experiments, which – after modification of the treatments and inclusion of subjects with different backgrounds – might yield more generalizable results in the future.

5 Conclusion

The empirical study presented in this paper investigated the effect of using one intelligent authoring shell, the xTEx-Sys system when compared to a learning content management system, Blackboard™. The effectiveness of the systems was analyzed by comparing the test results of students who used the xTEx-Sys system to the test results of students who used Blackboard™.

After the experiment results' analysis, we have calculated that the first checkpoint-test had a small effect size of 0.05 (there was no statistically significant difference between the groups) and the second check-point-test had no effect on learning outcomes.

Although we expected larger effect sizes, the results of the study were rather neutral. A possible explanation for the small, or no effect size, could be the fact that xTEx-Sys's domain knowledge presentation is somewhat technical and thus novel for an accustomed user. This issue was in the current experiment amplified by the fact that the knowledge database was created by a non-native English speaker outside US, while the system was used by native English speakers. This demonstrates the need of high language proficiency for the domain knowledge database designer.

As mentioned before, in order to further develop and improve the xTEx-Sys system, additional experiments should be conducted. The future experiments should determine options and venues for further improvement of the xTEx-Sys system that would yield greater student learning as well as greater user satisfaction with the system.

Acknowledgment

This work has been carried out within scientific project 177-0361994-1996 "Design and Evaluation of Intelligent E-learning Systems", funded by the Ministry of Science, Education and Sports of the Republic of Croatia.

References

[1] S. Stankov, A. Grubišić and B. Žitko, "E-learning paradigm & Intelligent tutoring systems", *Annual 2004 of the Croatian Academy of Engineering,* Z. Kniewald, ed., Croatian Academy of Engineering, Zagreb, pp. 21-31, 2004.

[2] J. Dempster, "Evaluating e-learning developments: An overview", warwick.ac.uk/go/cap/resources/eguides, retrieved 09.09.2008.

[3] D.H. Sleeman and J.S. Brown, "Introduction - Intelligent Systems", *Intelligent Tutoring Systems*, D.H. Sleeman and J. S. Brown, eds., Academic Press, Inc, London (Ltd), pp. 1-10, 1982.

[4] E. Wenger, "Artificial Intelligence and Tutoring Systems: Computational and Cognitive approaches to the Communication of Knowledge", Los Altos, California: Morgan Kaufmann, 1987.

[5] H.L. Burns and C.G. Capps, "Foundations of intelligent tutoring systems: an introduction", *Foundations of intelligent tutoring systems*, M. C. Poison and J. J. Richardson, eds., Lawrence Eribaum, London, pp. 1-19, 1988.

[6] P.A. Cohen, J.A. Kulik and C.L.C. Kulik, "Educational outcomes of tutoring: A meta-analysis of findings", *American Educational Research Journal* 19 (2), pp. 237-248, 1982.

[7] B.S. Bloom, "The Two-Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring", *Educational Researcher*. 13, pp. 4-16, 1984.

[8] G. Weber and P. Brusilovsky, "ELM-ART: An Adaptive Versatile System for Web-based Instruction", International Journal of Artificial Intelligence in Education, 12, pp. 351-384, 2001.

[9] Blackboard™, www.blackboard.com

[10] S. Stankov, Project TP-02/0177-01 Web oriented intelligent hypermedial authoring shell, Ministry of Science and Technology, Croatia, 2003-2005

[11] S. Stankov, "Isomorphic Model of the System as the Basis of Teaching Control Principles in an Intelligent Tutoring System", Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Croatia, PhD Thesis, 1997, (in Croatian)

[12] M. Rosić, "Establishing of Distance Education Systems within the Information Infrastructure", Faculty of Electrical Engineering and Computing, Zagreb, Croatia, MS Thesis, 2000, (in Croatian)

[13] S. Stankov, M. Rosic, B. Žitko and A. Grubišić, "TEx-Sys model for building intelligent tutoring systems", Computer & Education, Volume 51, Issue 3, pp. 1017-1036, 2008.

[14] A. Grubisic, "Evaluating Educational Influence of Intelligent E-learning Systems", Faculty of Electrical Engineering and Computing, Zagreb, Croatia, MS Thesis, 2007, (in Croatian)

[15] A. Iqbal, R. Oppermann, A. Patel and Kinshuk, "A Classification of Evaluation Methods for Intelligent Tutoring Systems", *Software Ergonomie '99*, U. Arend, E. Eberleh and K. Pitschke, eds., B. G. Teubner, Stuttgart, Leipzig, pp. 169-181, 1999.

[16] Fort Hays State University, FHSU PHYS 103 course, http://www.fhsu.edu/~zhrepic/Teaching/Courses/PHYS103/expSpring_PHYS103.htm, retrieved 09.09.2008.

[17] G.V. Glass, B. McGaw and M.L. Smith, "Meta-Analysis in Social Research", London: Sage, 1981.

[18] J. Cohen, "Statistical Power Analysis for the Behavioral Sciences", Hillsdale, NJ: Lawrence Erlbaum Ass., 1988.

Ani Grubišić is a scientific novice with a master's theses who works on the scientific project related to computational and didactical aspects of intelligent authoring tools in education and Web oriented intelligent hypermedia authoring shell. Her scientific interest is related to the e-learning systems' educational influence evaluation. She has authored or co-authored about ten scientific papers about the student modeling and effectiveness evaluation.

Slavomir Stankov is a full professor of computer science. Since 1996 he has been heading three scientific and one technological project. He has participated on some projects related to the information and communication technology applications in education, in particular the intelligent tutoring systems and authoring shells for their design, implementation and deployment. He has authored or co-authored more than seventy scientific and expert papers. He is a member of the Croatian Systems Society.

Zdeslav Hrepic is an assistant professor of physics at Fort Hays State University, KS, USA. He currently teaches range of introductory physics and astronomy courses and upper-level science methods courses. Between 1998 and now he taught variety of physics, polytechnics and information science courses at levels ranging from middle school (5th grade) to college. His research interests include educational technology, cognitive Issues in learning, development of formative assessment of students' conceptual understanding, issues in student understanding of traditional lecture, development of theoretical framework for analysis of interview data and transfer of learning.